Lectures in Mathematics ETH Zürich

Wolfgang Bangerth
Rolf Rannacher

Adaptive Finite Element Methods for Differential Equations

Birkhäuser

Wolfgang Bangerth and Rolf Rannacher
Adaptive Finite Element Methods for Differential Equations

The present lecture notes discuss concepts of "self-adaptivity" in the numerical solution of differential equations, with emphasis on Galerkin finite element methods. The key issues are *a posteriori* error estimation and *automatic* mesh adaptation. Besides the traditional approach of energy-norm error control, a new duality-based technique, the *Dual Weighted Residual* method for goal-oriented error estimation, is discussed in detail. This method aims at economical computation of arbitrary quantities of physical interest by properly adapting the computational mesh. This is typically required in the design cycles of technical applications. For example, the drag coefficient of a body immersed in a viscous flow is computed, then it is minimized by varying certain control parameters, and finally the stability of the resulting flow is investigated by solving an eigenvalue problem. "Goal-oriented" adaptivity is designed to achieve these tasks with minimal cost.

At the end of each chapter some exercises are posed in order to assist the interested reader in better understanding the concepts presented. Solutions and accompanying remarks are given in the Appendix. For the practical exercises, sample programs are provided via internet.

ISBN 3-7643-7009-2

www.birkhauser.ch

Lectures in Mathematics
ETH Zürich
Department of Mathematics
Research Institute of Mathematics

Managing Editor:
Michael Struwe

Wolfgang Bangerth
Rolf Rannacher

Adaptive Finite Element Methods for Differential Equations

Birkhäuser Verlag
Basel · Boston · Berlin

Authors' addresses:

Wolfgang Bangerth
TICAM
The University of Texas at Austin
201 E. 24th Street
Austin, TX 78712
USA
e-mail: bangerth@ticam.utexas.edu

Rolf Rannacher
Institute of Applied Mathematics
University of Heidelberg
Im Neuenheimer Feld 293
69120 Heidelberg
Germany
e-mail: rannacher@iwr.uni-heidelberg.de

2000 Mathematical Subject Classification 65L60, 65L70, 65M60, 65Nxx, 74S05, 76M10

A CIP catalogue record for this book is available from the
Library of Congress, Washington D.C., USA

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN 3-7643-7009-2 Birkhäuser Verlag, Basel – Boston – Berlin

ISBN 3-7643-7009-2

9 8 7 6 5 4 3 2 1

www.birkhauser.ch

Contents

Preface

These Lecture Notes have been compiled from the material presented by the second author in a lecture series ('Nachdiplomvorlesung') at the Department of Mathematics of the ETH Zürich during the summer term 2002. Concepts of 'self-adaptivity' in the numerical solution of differential equations are discussed with emphasis on Galerkin finite element methods. The key issues are *a posteriori* error estimation and *automatic* mesh adaptation. Besides the traditional approach of energy-norm error control, a new duality-based technique, the *Dual Weighted Residual* method (or shortly *DWR* method) for goal-oriented error estimation is discussed in detail. This method aims at economical computation of arbitrary quantities of physical interest by properly adapting the computational mesh. This is typically required in the design cycles of technical applications. For example, the drag coefficient of a body immersed in a viscous flow is computed, then it is minimized by varying certain control parameters, and finally the stability of the resulting flow is investigated by solving an eigenvalue problem. 'Goal-oriented' adaptivity is designed to achieve these tasks with minimal cost.

The basics of the DWR method and various of its applications are described in the following survey articles:

R. Rannacher [114], *Error control in finite element computations.* In: Proc. of Summer School Error Control and Adaptivity in Scientific Computing (H. Bulgak and C. Zenger, eds), pp. 247–278. Kluwer Academic Publishers, 1998.

M. Braack and R. Rannacher [42], *Adaptive finite element methods for low-Mach-number flows with chemical reactions.* In: 30th Computational Fluid Dynamics (H. Deconinck, ed.), Vol. 1999-03 of Lecture Series, von Karman Institute for Fluid Dynamics, Brussels, 1999.

R. Becker and R. Rannacher [31], *An optimal control approach to error estimation and mesh adaptation in finite element methods.* In: **Acta Numerica** 2000 (A. Iserles, ed.), pp. 1–101, Cambridge University Press, 2001.

R. Rannacher [117], *Duality techniques for error estimation and mesh adaptation in finite element methods.* In: Adaptive Finite Elements in Linear and Nonlinear Solid and Structural Mechanics (E. Stein, ed.), Vol. 416 of CISM Courses and Lectures, Springer, 2002, to appear.

R. Rannacher and F.-T. Suttmeier [121]. *Error estimation and adaptive mesh design for FE models in elasto-plasticity.* In: Error-Controlled Adaptive FEMs in Solid Mechanics (E. Stein, ed.), John Wiley, 2002, to appear.

Much of the contents of these Lecture Notes is taken from the above articles but new material has also been added. At the end of each chapter some exercises are posed in order to assist the interested reader in better understanding the presented concepts. Solutions and accompanying remarks are given in the Appendix. For these practical exercises, sample programs are provided at:
http://gaia.iwr.uni-heidelberg.de/httpdoc/Research/software.AFEMforDE.html.

Chapter 1

Introduction

We begin with a brief introduction to the philosophy underlying the approach to self-adaptivity which will be discussed in these Lecture Notes. Let the goal of a simulation be the accurate and efficient computation of the value of a functional $J(u)$, the 'target quantity', with accuracy TOL from the solution u of a continuous model by using an approximative discrete model of dimension N:

$$\mathcal{A}(u) = 0, \qquad \mathcal{A}_h(u_h) = 0.$$

The evaluation of the solution by the functional $J(\cdot)$ represents what exactly we want to know of a solution. Then, the goal of adaptivity is the 'optimal' use of computing resources according to either one of the following principles:

- Minimal work N for prescribed accuracy TOL,

$$N \to \min, \quad TOL \text{ given.}$$

- Maximal accuracy for prescribed work,

$$TOL \to \min, \quad N \text{ given.}$$

These goals are traditionally approached by automatic mesh adaptation on the basis of local 'error indicators' taken from the computed solution, assuming that this can indicate local roughness of the 'continuous' solution. The main ingredients of this process are:

- rigorous a posteriori error estimates in terms of data and the computed solution employing information about the continuous problem;
- local error indicators extracted from the a posteriori error estimates;
- automatic mesh adaptation according to certain refinement strategies based on the local error indicators.

We will demonstrate by examples that the appropriate choice of each of these steps is crucial for an economical simulation. Inappropriate realizations which violate the characteristic features of the underlying problem may drastically reduce the efficiency and accuracy.

The traditional approach to adaptivity aims at estimating the error with respect to the generic energy norm of the problem, or the global L^2-norm. However this is generally not what applications need. In the following, we will present a collection of examples for such situations, where one is really interested in computing locally defined quantities.

1.1 A first example: computation of drag coefficient

In order to illustrate the role of adaptivity in the design of a computational mesh, we consider a viscous incompressible flow around a cylinder in a channel with a narrowed outlet as shown in Figure 1.1. The flow quantities, velocity v and pressure p, are determined by the classical 'incompressible' Navier-Stokes equations

$$\partial_t v - \nu\Delta v + v\cdot\nabla v + \nabla p = f, \quad \nabla\cdot v = 0.$$

The configuration is two-dimensional, with Poisseuille inflow, and Reynolds number $\mathrm{Re} = 50$, such that the flow is laminar and stationary. The narrowing of the outlet causes a so-called *corner singularity* of the pressure.

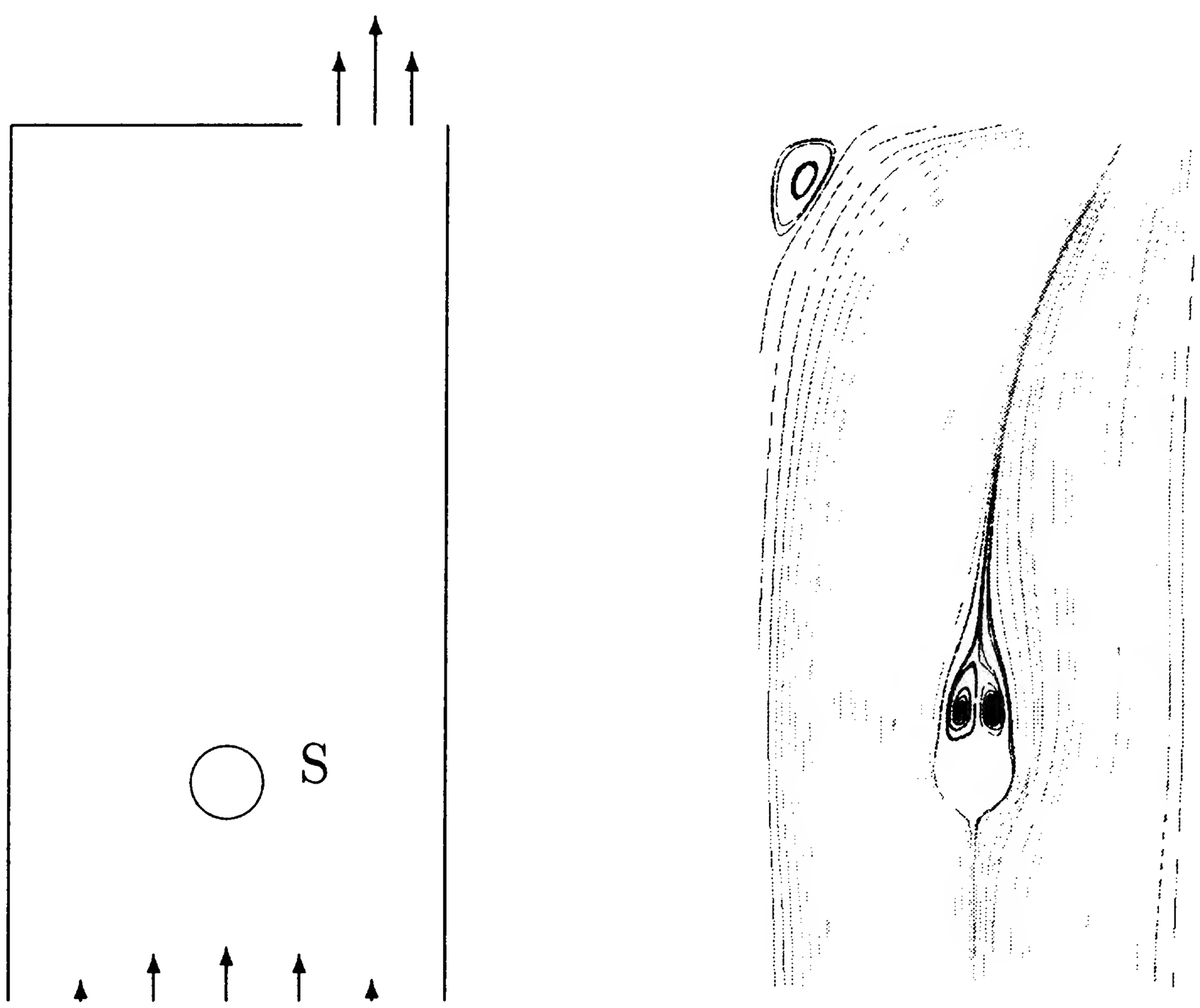


Figure 1.1: *Configuration and streamline plot for flow around a cylinder (*$\mathrm{Re} = 50$*).*

The goal is the accurate computation of the corresponding drag coefficient

$$J(v,p) := c_{\text{drag}} := \frac{2}{\bar{U}^2 D} \int_S n^T (2\nu\tau - pI) d \, \text{ds},$$

of the obstacle, where S is the surface of the body, D its diameter, $\bar{U}$ the maximal inflow velocity, $\tau := \frac{1}{2}(\nabla v + \nabla v^T)$ the strain tensor, and $d = (0,1)^T$ the main flow direction.

In order to control the mesh adaptation process in this simulation, one may find good reasons to use either of the following heuristic refinement indicators η_K on the mesh cells K:

- *Vorticity:* $\eta_K := h_K \|\nabla \times v_h\|_K$.
- *First-order pressure gradient:* $\eta_K := h_K \|\nabla p_h\|_K$.
- *Second-order velocity gradient:* $\eta_K := h_K \|\nabla_h^2 v_h\|_K$.
- *Residual-based indicator:* $\eta_K := h_K \|R_h\|_K + h_K^{1/2} \|r_h\|_{\partial K} + h_K \|\nabla \cdot v_h\|_K$,

$$R_{h|K} := f + \nu \Delta v_h - v_h \cdot \nabla v_h - \nabla p_h,$$

$$r_{h|\Gamma} := \left\{ \begin{array}{ll} \frac{1}{2}[\nu \partial_n v_h - n p_h], & \text{if } \Gamma \not\subset \partial\Omega \\ 0, & \text{if } \Gamma \subset \Gamma_{\text{rigid}} \cup \Gamma_{\text{in}} \\ -\nu \partial_n v_h + n p_h, & \text{if } \Gamma \subset \Gamma_{\text{out}} \end{array} \right\},$$

where n is the normal unit vector and $[\cdot]$ the jump across cell interfaces.

The vorticity as well as the pressure and velocity gradient indicators measure the 'smoothness' of $\{v_h, p_h\}$, while the heuristical residual-based indicator additionally contains information about local conservation of momentum and mass. As competitors, we additionally consider global uniform refinement and refinement using a new approach, called *DWR method* (Dual Weighted Residual method) which uses the same residual terms as in the heuristic residual-based indicators but multiplied by weights obtained by solving a global 'dual' problem. This method will be systematically developed in these Lecture Notes.

The above test case has been designed in order to demonstrate the ability of the different refinement indicators to produce meshes on which the main features of the flow, such as boundary layers along rigid walls, vortices behind the cylinder, and the corner singularity at the outlet, are sufficiently resolved. Figure 1.2 shows locally adapted meshes obtained on the basis of the different refinement indicators. The results shown in Figure 1.3 demonstrate that in this case the two ad hoc indicators involving only the norm of vorticity or pressure gradient are even less efficient than simple uniform refinement. This demonstrates that a systematic approach to goal-oriented mesh adaptation is needed which not only takes into account local properties of the solution but also the global dependence of the error in the target quantity on these properties.

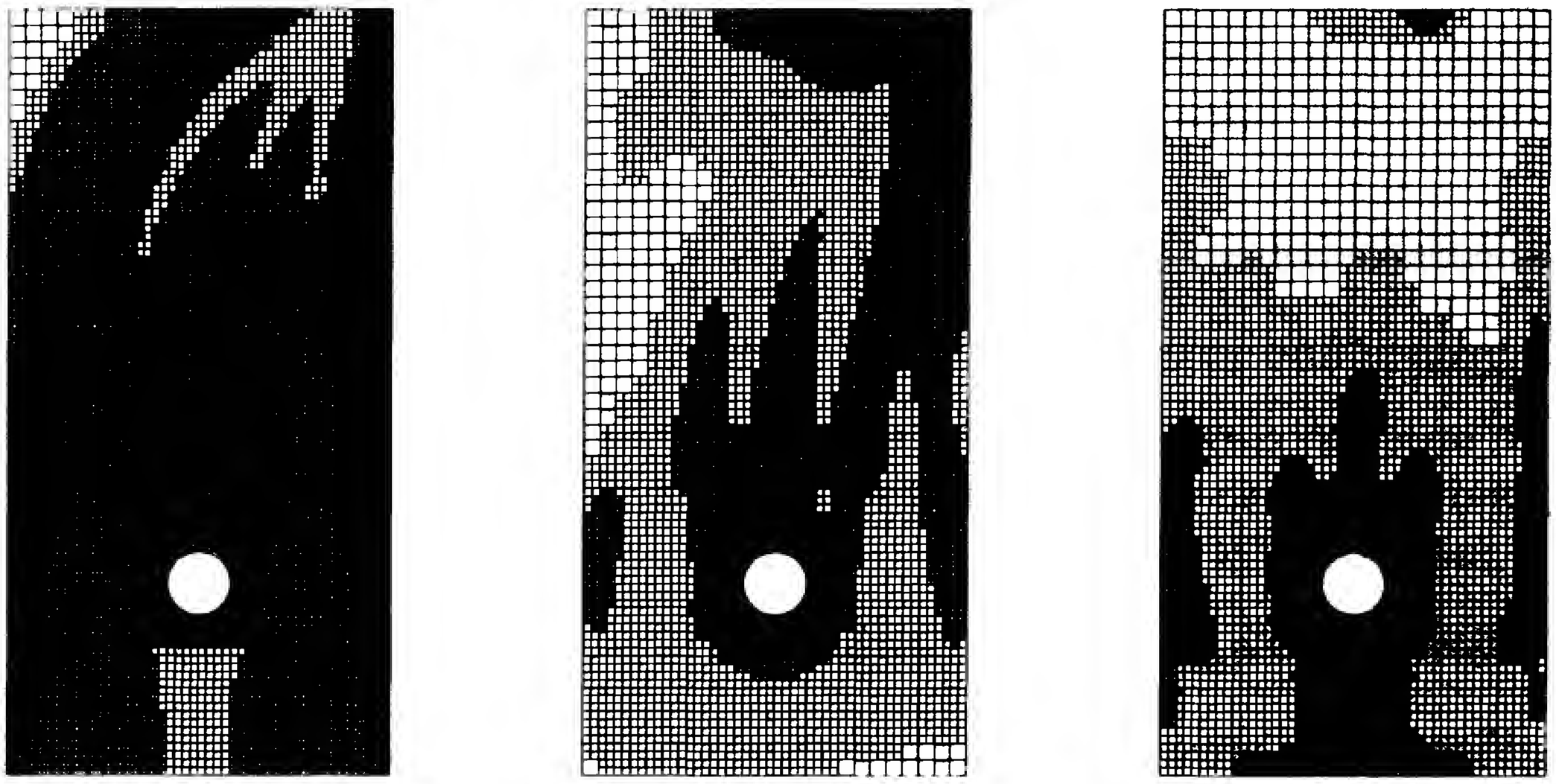


Figure 1.2: *Meshes with 5,000 cells obtained by the vorticity indicator (left), the heuristic residual-based indicator (middle), and the new DWR indicator (right).*

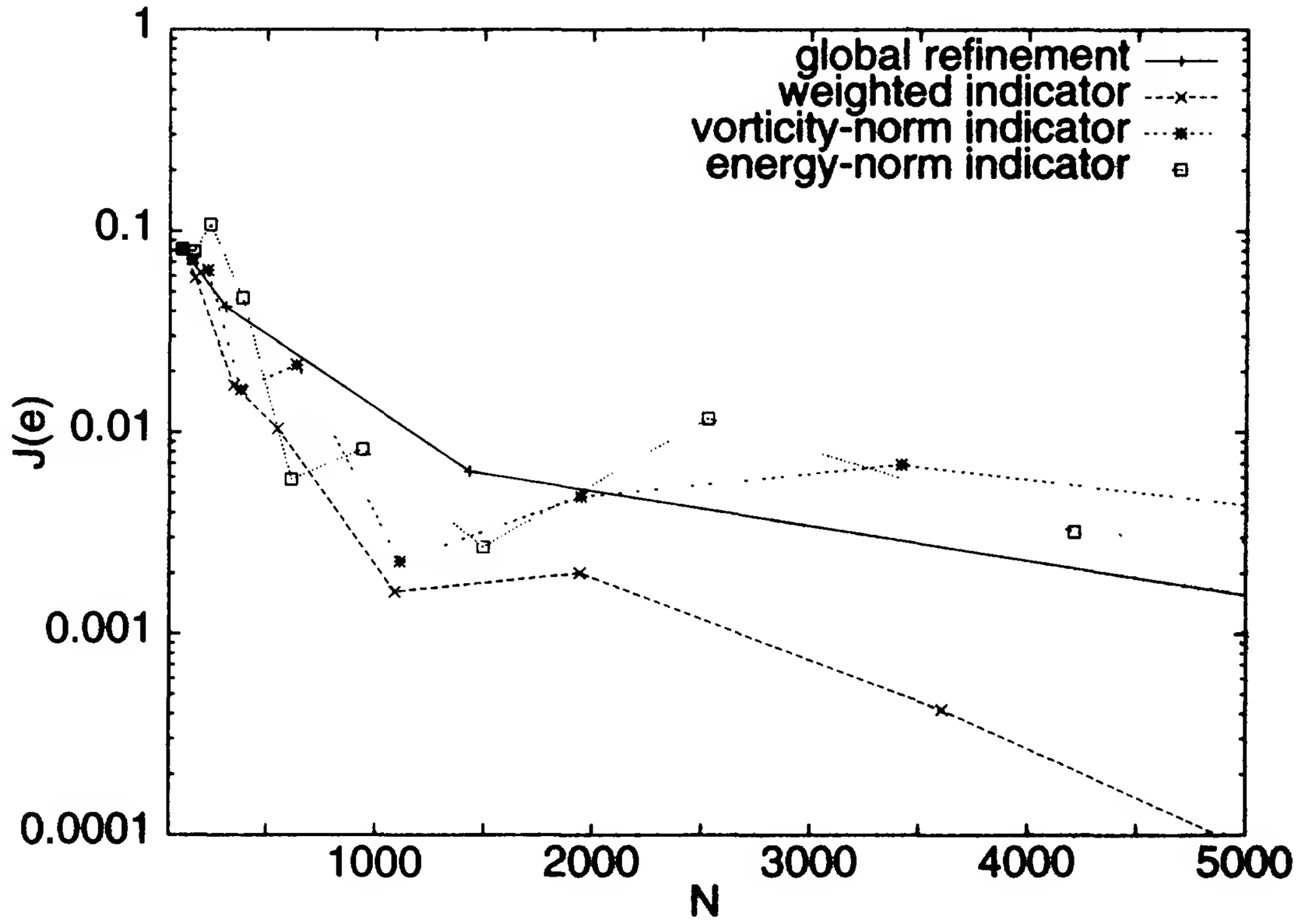


Figure 1.3: *Error $J(e)$ in the drag coefficient versus number of cells N, for uniform refinement, the weighted indicator obtained by the DWR approach, the vorticity indicator, and the heuristic residual-based indicator.*

1.2 The need for 'goal-oriented' mesh adaptation

Let us illustrate the need for 'goal-oriented' mesh adaptation by two further examples of different types of complexity: a 3-d flow problem where only on locally

adapted meshes sufficient accuracy is achieved with acceptable costs, and a 2-d flow problem in which the complication arises through the strong interaction of mass transport and heat transfer. In both situations, the main problem is the generation of error estimates which reflect the local and global dependency of the error on the locally observed solution properties. In fact, usually mesh adaptation in solving a coupled system of equations for a set of physical quantities $u = (u_h^i)_{i=1}^n$ is based on 'smoothness' or 'residual' information like

$$\eta_K := \sum_{i=1}^{n} \omega_K^i \| D_h^2 u_h^i \|_K \qquad \text{or} \qquad \eta_K = \sum_{i=1}^{n} \omega_K^i \| R_i(u_h) \|_K .$$

Here, $D_h^2 u_h^i$ stands for certain second-order difference quotients, and $R_i(u_h)$ are certain residuals as introduced in the previous example. Both kinds of indicators are easily evaluated from the computed solution and are widely used in practice. The proper choice of the weights ω_K^i is crucial for the effectivity of the adaptation process. They should include both a scaling due to different orders of magnitude of the solution components u^i, as well as the influence of the present cell K on the requested quantity of interest.

A cylinder flow benchmark in 3-D

We consider the 3-dimensional flow in a channel around a cylinder with square cross section as shown in Figure 1.4. The Reynolds number is $\mathbf{Re} = 20$, such that the flow is laminar and stationary. This is part of a benchmark suite for the computation of viscous, incompressible fluid flow (see Schäfer and Turek [124]).

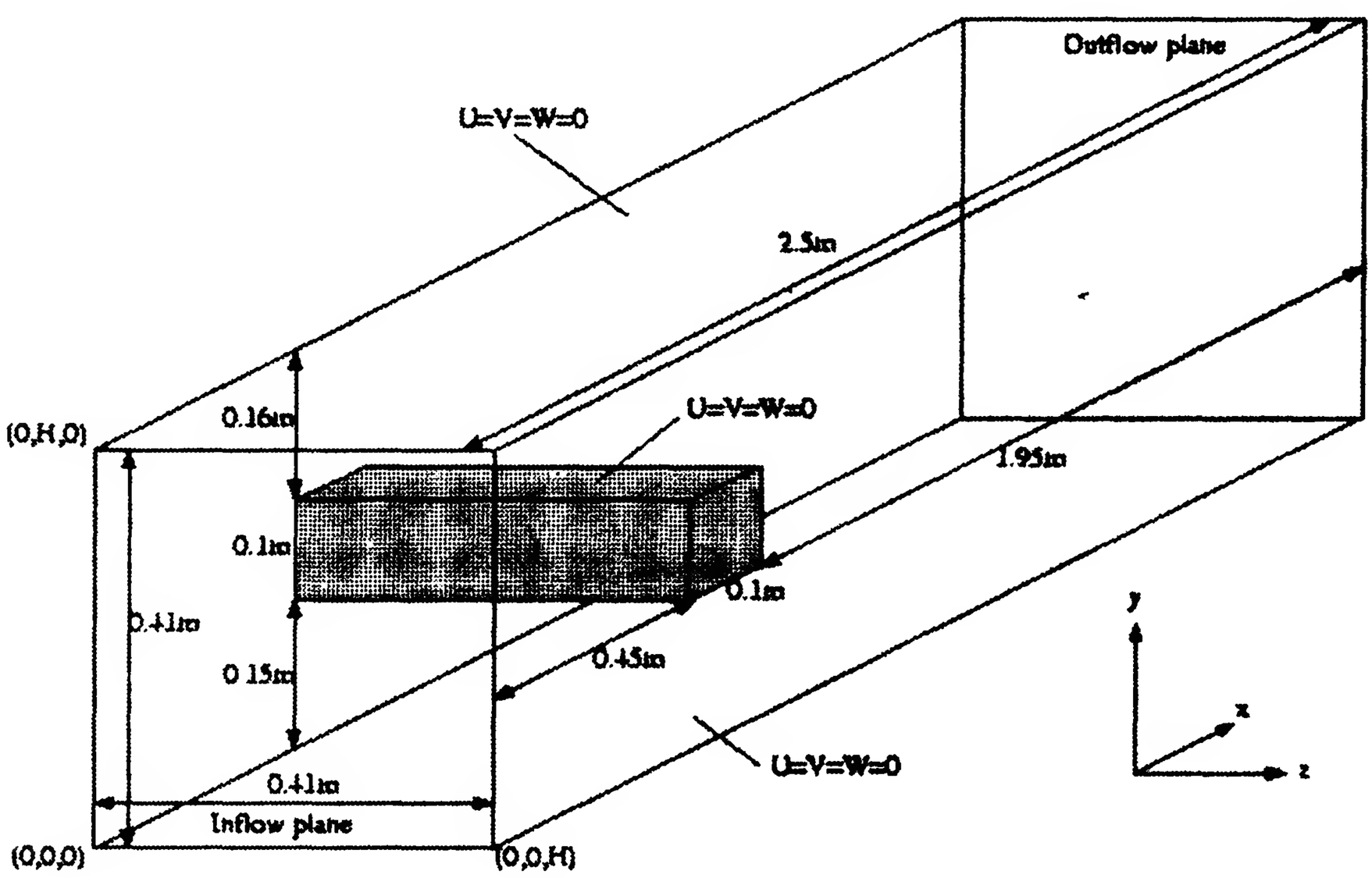


Figure 1.4: *Configuration of 3D flow around a square cylinder in a channel.*

The goal is again the accurate computation of the **drag coefficient**

$$J(v,p) := c_d = \frac{2}{\bar{U}^2 DH} \int_S n^T (2\nu\tau - pI) d \, \mathrm{ds},$$

where D, H are geometrical parameters, $\bar{U}$ the maximum inflow **velocity and** $d = (1,0,0)^T$ the main flow direction. The desired accuracy is $\mathrm{TOL} \sim 1\%$ **which** turns out to be a rather demanding task even for this simple **flow situation.**

In Table 1.1, we compare the efficiency of global uniform **refinement against** that of local refinement on the basis of a heuristic **residual-based indicator and** the DWR method as already used in the first example. **The superiority of mesh** adaptation by the DWR method is clearly seen.

Table 1.1: *Drag results: a) Q_2/Q_1-element with global refinement, b) Q_1/Q_1-element with local refinement by heuristic residual indicator, c) Q_1/Q_1-element with local refinement by DWR method; the first mesh level on which an error smaller than 1% is achieved is indicated in boldface; from Braack et al. [40].*

a) N	c_d	b) N	c_d	c) N	c_d
15,960	8.2559	3,696	12.7888	3,696	12.7888
117,360	7.9766	21,512	8.7117	8,456	9.8262
899,040	7.8644	80,864	7.9505	15,768	8.1147
7,035,840	7.8193	182,352	7.9142	30,224	8.1848
55,666,560	7.7959	473,000	7.8635	**84,832**	7.8282
–	–	**1,052,000**	7.7971	162,680	7.7788
∞	7.7730	∞	7.7730	∞	7.7730

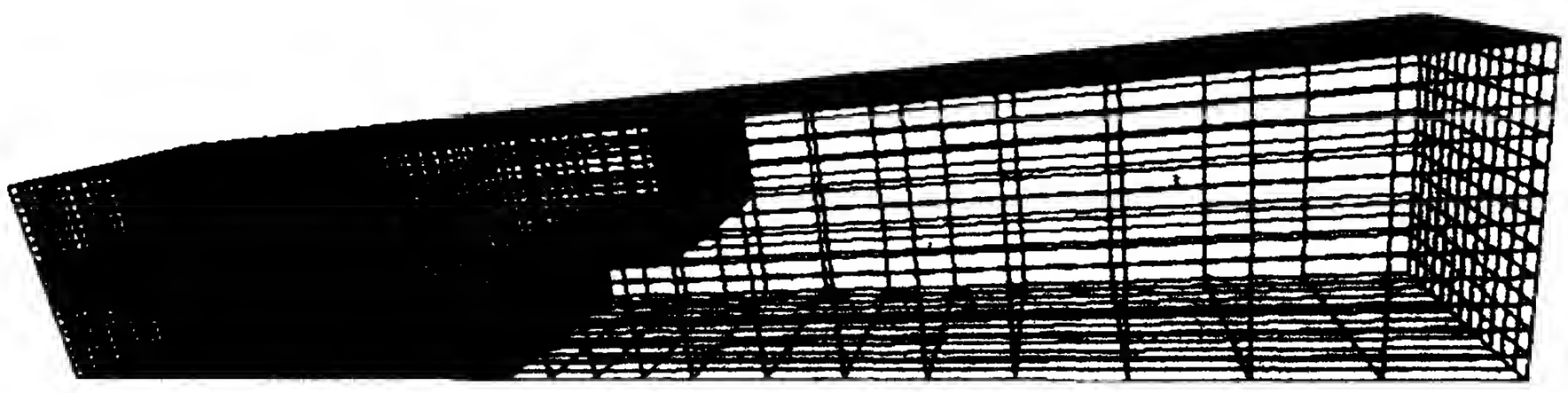


Figure 1.5: *Refined mesh and zoom into the cylinder vicinity obtained by the heuristic residual-based indicator; from Braack et al. [40].*

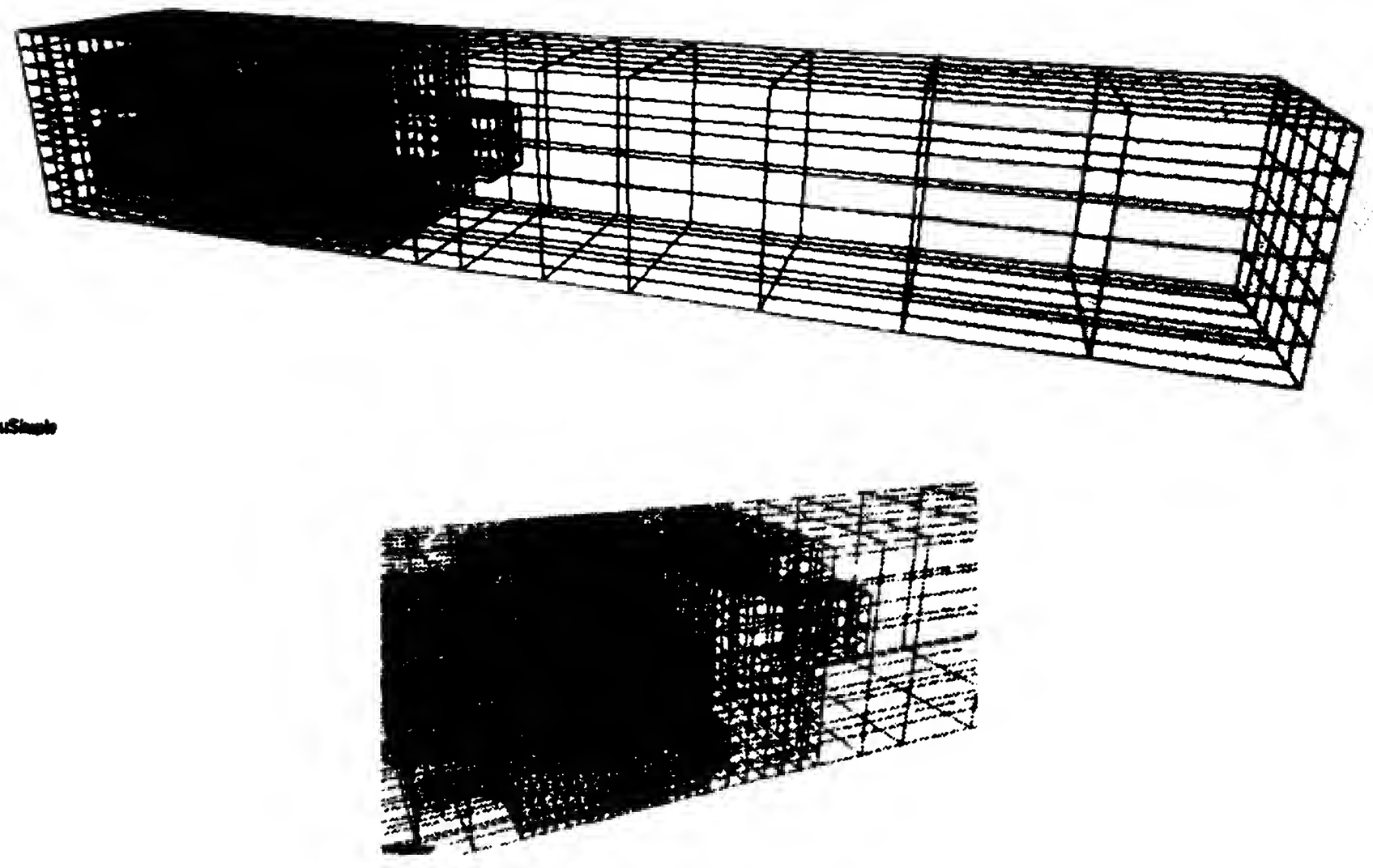


Figure 1.6: *Refined mesh and zoom into the cylinder vicinity obtained by the DWR method; from Braack et al. [40].*

Figures 1.5 and 1.6 show meshes which have been obtained by using refinement on the basis of the heuristic residual-based indicator and by the DWR method. The heuristic indicator fails to properly refine the area behind the cylinder which causes its poorer drag approximation.

A heat-driven cavity benchmark in 2-D

We consider a 2-dimensional cavity flow. The flow in a square box with side length $L = 1$ (see Figure 1.7) is driven by a temperature difference $\theta_h - \theta_c = 720\,K$, between the left ('hot') and the right ('cold') wall, under the action of gravity g in the vertical direction. The Rayleigh number is $Ra \sim 10^6$ making this problem computationally demanding. Here, the quantity to be computed is the average Nusselt number (mean heat flux) along the cold wall defined by

$$J(u) = \langle \mathrm{Nu} \rangle_c := \frac{\mathrm{Pr}}{0.3\mu_0\theta_0} \int_{\Gamma_{cold}} \kappa \partial_n \theta \, ds,$$

where Pr is the Prandtl number and μ_0, θ_0 are certain reference values for viscosity and temperature. The underlying mathematical model is the *low-Mach number approximation* of the stationary compressible Navier-Stokes equations which is expressed in terms of the set of primitive variables $u = \{v, p, \theta\}$ denoting velocity, pressure and temperature. In this case, due to the large temperature difference, the usual Boussinesq approximation is not sufficient (see **Becker and Braack** [24]).

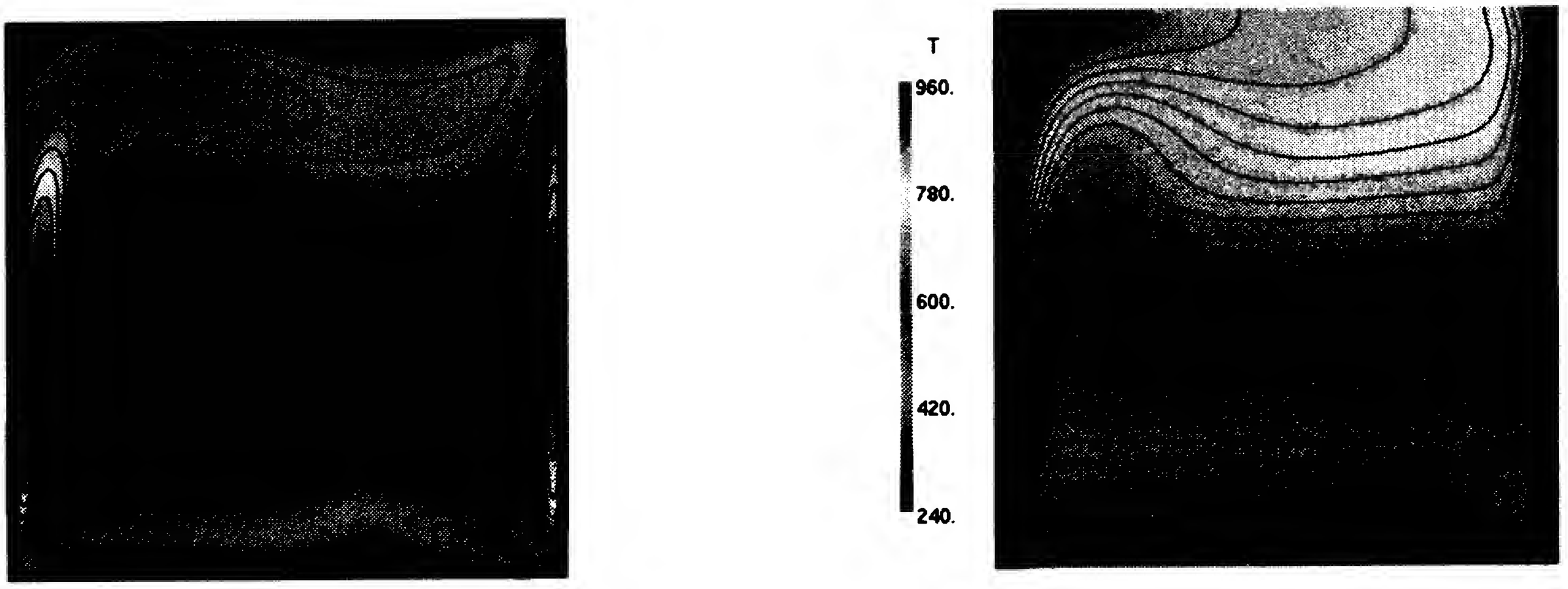


Figure 1.7: *Flow in heat-driven cavity: velocity norm isolines (left) and temperature isolines (right); from Becker and Braack [24].*

The meshes shown in Figure 1.8 indicate that the heuristic residual-based indicator induces mesh refinement mainly in those areas where the velocity is dominant while the weighted error indicator obtained by the DWR method puts more emphasis on the region along the hot boundary where the temperature gradient is dominant. The latter, given the quantity we want to compute, seems to be more important for capturing the heat transfer through the cavity. This is confirmed by the results shown in Table 1.2 which show that on the meshes generated by the properly weighted error indicators the accuracy in computing the Nusselt number is better by almost one order.

Table 1.2: *Computation of the Nusselt number in the heat-driven cavity by the heuristic residual-based indicator (left) and the weighted indicator by the DWR method (right); comparable error magnitudes are indicated in boldface; from Becker and Braack [24].*

N	$\langle Nu \rangle_c$	error
524	-9.09552	$4.1 \cdot 10^{-1}$
945	-8.67201	$1.5 \cdot 10^{-2}$
1708	-8.49286	$1.9 \cdot 10^{-1}$
3108	-8.58359	$1.0 \cdot 10^{-1}$
5656	-8.59982	$8.7 \cdot 10^{-2}$
18204	-8.64775	$3.9 \cdot 10^{-2}$
32676	-8.66867	$1.8 \cdot 10^{-2}$
58678	-8.67791	$8.7 \cdot 10^{-3}$
79292	-8.67922	$7.4 \cdot 10^{-3}$

N	$\langle Nu \rangle_c$	error
523	-8.86487	$1.8 \cdot 10^{-1}$
945	-8.71941	$3.3 \cdot 10^{-2}$
1717	-8.66898	$1.8 \cdot 10^{-2}$
5530	-8.67477	$1.2 \cdot 10^{-2}$
9728	-8.68364	$3.0 \cdot 10^{-3}$
17319	-8.68744	$8.5 \cdot 10^{-4}$
31466	-8.68653	$6.9 \cdot 10^{-5}$

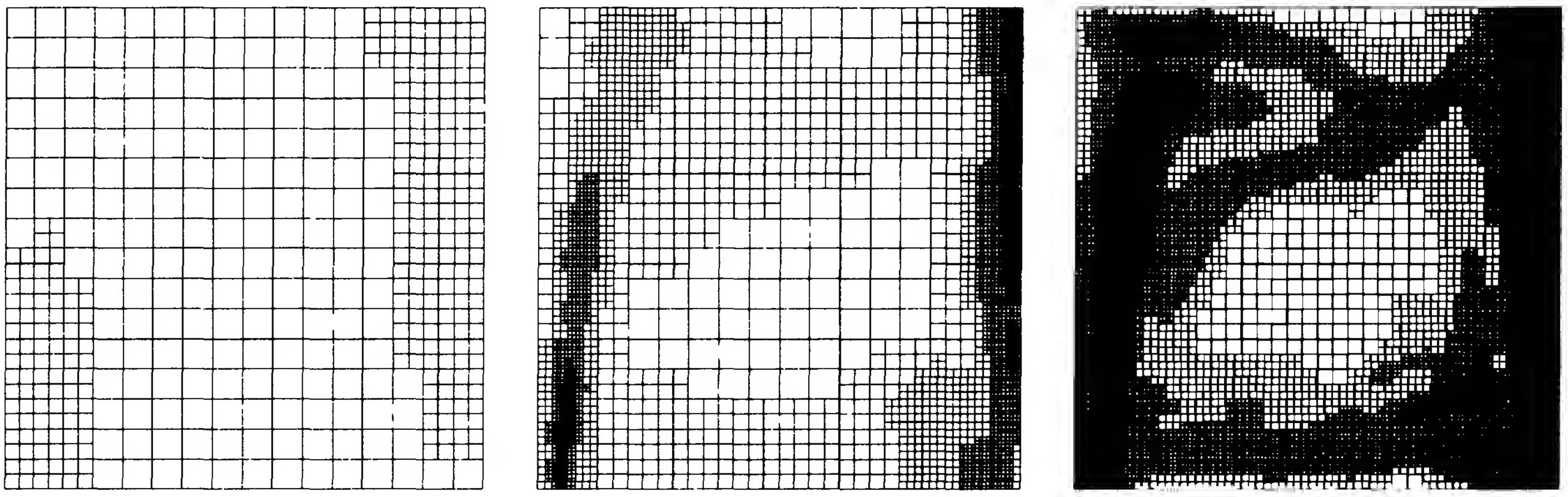


Figure 1.8: *Sequence of refined meshes for the heat-driven cavity with* $N \approx 500, 5500, 56000$ *cells: heuristic residual-based indicator (top row), weighted error indicator by the DWR method (bottom row); from Becker and Braack [24].*

1.3 Further examples of goal-oriented simulation

In this section, let us present a collection of further examples from different areas in which 'goal-oriented' adaptivity is needed and has already proven to be superior over other more ad-hoc approaches. Some of these examples will be discussed in more detail in the course of these Lecture Notes.

Example 1.1. *CARS signal in a flow reactor* (Becker et al. [26]):

$$J(p, v, \theta, w) := \kappa \int_{\Gamma_{CARS}} w_i^2 \, \sigma \, do.$$

In chemical flow reactors **C**oherent **A**ntistokes **R**aman **S**pectroscopy (CARS) is used to produce signals which reflect the distributions of the mole fractions of certain reaction products w_i along the measurement line Γ_{CARS}. The mole fraction of the species w_i is determined by the balance equation

$$c_p \partial_t w_i - \nabla(\rho D_i \nabla w_i) + \rho v \cdot \nabla w_i = f_i(\theta, w),$$

together with equations for the other physical quantities, pressure p, velocity v, temperature θ, and density ρ. The final goal is to determine certain reaction

velocities in the chemical process from these measurements. The accurate computation of the corresponding quantities is of crucial importance for this parameter estimation process.

Example 1.2. *Mean surface pressure of a body in an inviscid flow* (Hartmann [71]):

$$J(\rho, v, e) := \int_S p \, do.$$

In the absence of viscosity, this quantity relates to the drag coefficient of the body. The density ρ, the momentum ρv and the energy e are determined by the Euler equations, setting $p = (\gamma - 1)(e - \frac{1}{2}\rho v^2)$,

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho v) &= 0, \\ \partial_t (\rho v) + \nabla \cdot (\rho v \otimes v) + \nabla p &= \rho g, \\ c_p \partial_t (\rho e) + c_p \nabla \cdot (\rho e v + p v) &= h. \end{aligned}$$

Example 1.3. *Boundary mean stress of an elasto-plastic body* (Rannacher and Suttmeier [119]),

$$J(u, \sigma) := \int_{\Gamma_D} n^T \sigma n \, do,$$

over the 'clamped' part Γ_D of the boundary. The stress tensor σ and displacement u are determined by the Lamé-Navier equations with a pointwise stress constraint,

$$-\nabla \cdot \sigma = f, \quad \sigma = C\varepsilon, \quad \varepsilon = \tfrac{1}{2}(\nabla u + \nabla u^T), \quad |\sigma| \le \sigma_0.$$

Example 1.4. *Local intensity measurement of a seismic signal* (Bangerth [14]):

$$J(u) := \int_{T-\delta}^{T+\delta} u(x_{\text{obs}}, t) \, dt.$$

The displacement u is determined by the acoustic or elastic wave equation

$$\rho \partial_t^2 u - \nabla \cdot (a \nabla u) = 0.$$

In applications the elastic coefficient a has to be determined from measurements of this kind at varying points x_{obs} resulting in an inverse problem.

Example 1.5. *Observed light emission of a proto-stellar dust cloud* (Kanschat [93]):

$$J(u) := \int_{n \cdot \theta_{obs} \ge 0} u(x, \theta_{\text{obs}}, \lambda_0) \, do.$$

The radiation intensity u is determined by the *radiative transfer equation*

$$r_\theta \cdot \nabla u + (\kappa + \mu) u = B - \int_{S^d} R(\theta, \theta') u \, d\theta'.$$

Due to the distance of the light-emitting object, the observer (located on a satellite) measures only the mean value of the intensity integrated over that part of the surface of the object seen by the observer, and at some fixed wave length λ_0.

Example 1.6. *Critical eigenvalue in hydrodynamic stability analysis* (Heuveline and Rannacher [77]):

$$J(p, v, \lambda) := \lambda^{crit}.$$

The triple $\{p, v, \lambda\}$ is determined by the eigenvalue problem of the Navier-Stokes equations linearized about some stationary base flow $\hat{v}$,

$$-\nu\Delta v + \hat{v}\cdot\nabla v + v\cdot\nabla\hat{v} + \nabla p = \lambda v, \quad \nabla\cdot v = 0.$$

Example 1.7. *Cost functional in an optimization problem* (Becker et al. [28]):

$$J(u, q) := J_{\text{cost}}(u, q).$$

The state variable v and control variable q are determined by

$$J_{\text{cost}}(u, q) \to \min, \quad \mathcal{A}(u) + Bq = 0.$$

The state equation may, for example, consist of the 'incompressible' Navier-Stokes equations containing some control parameter q (e.g., boundary control) and the cost functional may be the drag coefficient which is to be minimized.

1.4 General concepts of error estimation

In the following, we will introduce some of the main concepts of the DWR method for 'goal-oriented' a posteriori error estimation within the framework of linear algebra. We will use the same concepts later for differential equations as well and use this simple example only to introduce the most important aspects.

Traditional error estimation

For regular matrices A, $A_h \in \mathbb{R}^{n\times n}$, and vectors b, $b_h \in \mathbb{R}^n$, consider the problems of finding $x, x_h \in \mathbb{R}^n$ from

$$Ax = b, \qquad A_h x_h = b_h, \tag{1.1}$$

where h is a parameter indicating the quality of approximation, i.e., $A_h \to A$ and $b_h \to b$, as $h \to 0$. In this context, we introduce the notation of the *approximation error* $e := x - x_h$, the *truncation error* $\tau := A_h x - b_h$, and the *residual* $\rho := b - Ax_h$. Usually, *a priori* error analysis is based on the truncation error, and uses the identity

$$A_h e = A_h x - A_h x_h = A_h x - b_h = \tau,$$

to derive an a priori error bound involving a 'discrete' stability constant $c_{S,h}$:

$$\|e\| \le c_{S,h}\,\|\tau\|, \qquad c_{S,h} := \|A_h^{-1}\|. \tag{1.2}$$

In contrast to that, the *a posteriori* error analysis uses the relation

$$Ae = Ax - Ax_h = b - Ax_h = \rho,$$

to derive an a posteriori error bound involving a 'continuous' stability constant:

$$\|e\| \leq c_S \, \|\rho\|, \qquad c_S := \|A^{-1}\|. \tag{1.3}$$

Notice that the *a priori* error analysis is based on assumptions on the stability properties of the 'discrete' operator A_h, which may be difficult to establish for the particular approximation, while the *a posteriori* error analysis uses stability properties of the unperturbed 'continuous' operator A which are often available from regularity theory. Further, the truncation error τ is not so easily computable in practical applications.

Duality-based error estimation

In order to avoid the aforementioned drawbacks and to estimate the error also with respect to arbitrary moments of the solution, we employ a 'duality argument' well-known from the error analysis of Galerkin methods. For some given $j \in \mathbb{R}^n$ assume that we want to estimate the value of the linear error functional

$$J(e) = J(u) - J(x_h) = (e, j).$$

For the determination of this error, consider the solution $z \in \mathbb{R}^n$ of the associated *dual* (or *adjoint*) *problem*

$$A^* z = j. \tag{1.4}$$

This leads us to an identity about the error

$$J(e) = (e, j) = (e, A^* z) = (Ae, z) = (\rho, z),$$

and finally to the 'weighted' a posteriori error estimate

$$|J(e)| \leq \sum_{i=1}^{n} |\rho_i| \, |z_i|. \tag{1.5}$$

In this estimate the residuals ρ_i are easily computable but the determination of the weights z_i requires the solution of the auxiliary problem (1.4). The gain in using these weights is that they tell us about the influence of the 'local' residuals ρ_i on the error in the target quantity $J(u)$.

We want to extend the concept of duality-based a posteriori error analysis to nonlinear problems. Let differentiable mappings $A(\cdot)$, $A_h(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ and vectors b, $b_h \in \mathbb{R}^n$ be given and consider the problems

$$A(x) = b, \qquad A_h(x_h) = b_h. \tag{1.6}$$

Using the Jacobi matrix $A'(x)$ and the nonlinear residual $\rho := b - A(x_h)$, we have the following identities, for an arbitrary $y \in \mathbb{R}^n$:

$$(A(x)-A(x_h), y) = \int_0^1 (A'(x_h+se)e, y)\, ds = (Be, y),$$

with the matrix

$$B := B(x, x_h) := \int_0^1 A'(x_h+se)\, ds.$$

For any given error functional $J(\cdot) = (\cdot, j)$, let $z \in \mathbb{R}^n$ be the solution of the corresponding dual problem

$$B^* z = j,$$

where regularity of B^* is assumed. We obtain the error identity

$$J(e) = (e, j) = (e, B^* z) = (Be, z) = (A(x)-A(x_h), z) = (\rho, z),$$

which leads us to the weighted a posteriori error estimate

$$|J(e)| \le \eta := \sum_{i=1}^{n} |\rho_i|\,|z_i|. \tag{1.7}$$

The use of the above error estimate requires the evaluation of the weights $|z_i|$. However, z cannot be computed since it depends on the (unknown) error e through the definition of B. To this end, we approximate the matrix B by

$$B(x, x_h) \approx \tilde{B} := B(x_h, x_h) = \int_0^1 A'(x_h)\, ds = A'(x_h),$$

and solve the linearized dual problem, i.e. in practice an approximation of that,

$$A_h'(x_h)^* \tilde{z}_h = j_h.$$

This leads us to the approximate error estimate

$$|J(e)| \approx \tilde{\eta} := |(\rho, \tilde{z}_h)| \le \sum_{i=1}^{n} \tilde{\omega}_i |\rho_i|, \qquad \tilde{\omega}_i := |\tilde{z}_{h,i}|. \tag{1.8}$$

The error by this approximation can be estimated. With $\tilde{z}$ the solution of

$$\tilde{B}^* \tilde{z} = A'(x_h)^* \tilde{z} = j,$$

we have by definition:

$$\begin{aligned} |J(e)| &= |(e, \tilde{B}^* \tilde{z})| = |(\tilde{B}e, \tilde{z})| \\ &\le |((\tilde{B}-B)e, \tilde{z})| + |(Be, \tilde{z})| \\ &\le |((\tilde{B}-B)e, \tilde{z})| + |(\rho, \tilde{z})| \\ &\le |((\tilde{B}-B)e, \tilde{z})| + |(\rho, \tilde{z}-\tilde{z}_h)| + \tilde{\eta}. \end{aligned}$$

Using the estimate,

$$\|\tilde{B}-B\| = \left\| \int_0^1 \{A'(x_h)-A'(x_h+se)\}\, ds \right\| \leq \tfrac{1}{2}L'\|e\|,$$

with the Lipschitz constant L' of A', we obtain

$$|J(e)| \leq \tilde{\eta} + \tfrac{1}{2}L'\|e\|^2\|\tilde{z}\| + \|\rho\|\,\|\tilde{z}-\tilde{z}_h\|. \tag{1.9}$$

Hence, assuming that $\|\tilde{z}\|$ can be controlled, the approximate error estimator $\tilde{\eta}$ is the dominant error term. However, notice that the derivation of (1.9) is based on typical 'linear algebra' arguments as it does not observe a possible dimension-dependence of norms and constants. In the PDE context, considered later on, we will have to argue more carefully in estimating the effect of linearization and approximation of the dual problem.

Overview

The further contents of these Lecture Notes are as follows: In Chapter 2, we apply the principle of duality-based error estimation to the Galerkin approximation of ODEs. The following Chapter 3 is one of the core parts of these Lecture Notes. There, we develop the DWR method for the Poisson equation as the prototype of an elliptic problem. Then, Chapter 4 is devoted to the practical realization of error estimation and mesh adaptation. So far, the development of the DWR method is largely based on heuristic grounds, though strongly supported by computational experience. In Chapter 5, we discuss some of the central theoretical questions related to the justification of this approach but quickly get to the limits of theoretical analysis. Chapter 6 is the second core part of these Lecture Notes. It presents an abstract version of the DWR method for general nonlinear variational problems. This abstract approach is used in Chapter 7 for developing an a posteriori error analysis for the Galerkin approximation of eigenvalue problems. Another application is presented in Chapter 8, where the Galerkin approximation of optimization problems with PDE constraints is considered using the classical Lagrangian formalism. In Chapter 9, we realize the DWR method for the space-time discretization of nonstationary problems, with the heat and wave equations as model cases. Chapter 10 deals with the applications of the DWR method to linear and nonlinear problems from Structural Mechanics and Chapter 11 contains various results on the approximation of the incompressible Navier-Stokes equations as one of the basic models in fluid mechanics. In the last Chapter 12, some current developments and open problems are addressed.

Chapter 2

An ODE Model Case

In the following, we consider the realization of the ideas sketched in the Introduction for the initial value problem of an autonomous ODE system:

$$u'(t) = f(u(t)), \quad t \in I := [0, T], \quad u(0) = u_0. \tag{2.1}$$

We assume that the function $f(\cdot)$ is Lipschitz continuous and that the solution u exists on the interval $[0, T]$. Very often, in applications, the goal in numerically approximating the solution is to know the end-time value

$$J(u) = u(T).$$

This goal may be reached by using the traditional *Finite Difference* (FD) method or the Galerkin *Finite Element* (FE) method with suitable error control and step-size selection strategies. The material of this chapter is mainly taken from Böttcher [37], and Böttcher and Rannacher [38].

2.1 Finite differences and finite elements

We start with a brief outline of the main philosophies underlying the Finite Difference and the Galerkin Finite Element schemes. Both types of methods are defined on time grids on the interval I described by

$$0 = t_0 < \ldots < t_n < \ldots < t_N = T, \quad I_n = (t_{n-1}, t_n], \quad k_n = t_n - t_{n-1}.$$

The finite difference method

We exemplarily consider the (implicit) *backward Euler scheme*: Find $U_n \sim u_n := u(t_n)$, such that

$$U_0 = u_0, \qquad U_n = U_{n-1} + k_n f(U_n) \quad n \geq 1. \tag{2.2}$$

Assuming the step size k_n to be sufficiently small (or $f(\cdot)$ to be negative monotone), the discrete solution U_n exists for all $n = 1, \ldots, N$. The corresponding truncation error

$$\tau_n(u) := k_n^{-1}(u_n - u_{n-1}) - f(u_n),$$

is bounded like

$$\|\tau_n(u)\| \leq \tfrac{1}{2} k_n \max_{I_n} \|u''\|.$$

Viewing the error $e_n = u_n - U_n$ as particular solution of the discrete scheme with right-hand side τ_n, the discrete Gronwall lemma yields the usual a priori error estimate

$$\|e_N\| \leq K \sum_{n=1}^{N} k_n \|\tau_n(u)\|, \qquad K \sim \exp\Big(\int_0^T L_f(t)\, dt \Big), \tag{2.3}$$

where $L_f(t)$ is the Lipschitz constant of $f(\cdot)$ along the exact solution u. On this basis, we may use the following *explicit* formula for step-size control:

$$k_n = \frac{\mathrm{TOL}}{KT\|\tau_n^0(U)\|}, \tag{2.4}$$

where the following approximation of the truncation error is used:

$$\tau_n(u) = k_n \tau_n^0(u) + \mathcal{O}(k_n^2) \approx k_n \tau_n^0(U).$$

Remark 2.1. The *a priori* error estimate (2.3) suffers from two short-comings:

- Usually the growth factor $K = K(T, L_f)$ is unknown as it depends on the exact solution.
- The step size formula requires an estimate for the 'leading' truncation error $\tau_n^0(u)$ along the exact solution. This, however, has to be generated (for example by local h-extrapolation) from the computed solution U_n.

The Galerkin finite element method

We consider the approximation by the so-called *discontinuous Galerkin dG(0) (finite element) method.* Here, the space

$$S_h^{(0)}(I) := \{\varphi : I \to \mathbb{R}^d, \varphi_{|I_n} \in P_0(I_n)\},$$

consisting of piecewise constant functions, is used as 'trial' and 'test' space. The 'Galerkin approximation' $U \in S_h^{(0)}$ is determined by requiring that $U_0^- := u_0$ and

$$\sum_{n=1}^{N} \Big\{ \int_{I_n} (U' - f(U), \psi)\, dt + ([U]_{n-1}, \psi_{n-1}^+) \Big\} = 0 \quad \forall \psi \in S_h^{(0)}. \tag{2.5}$$

Here, we have used the standard notation

$$\varphi_n^- := \lim_{t\uparrow t_n} \varphi(t), \quad \varphi_n^+ := \lim_{t\downarrow t_n} \varphi(t), \quad [\varphi]_n := \varphi_n^+ - \varphi_n^-,$$

for left- and right-sided limits, and jumps of possibly discontinuous functions. Clearly, the continuous solution u also satisfies the variational equation (2.5), and for the error $e := u - U$, there holds the nonlinear *Galerkin orthogonality* relation

$$\sum_{n=1}^{N} \Big\{ \int_{I_n} (e' - f(u) + f(U), \psi)\, dt + ([e]_{n-1}, \psi_{n-1}^+) \Big\} = 0 \quad \forall \psi \in S_h^{(0)}. \tag{2.6}$$

Since the test functions are allowed to be discontinuous, the globally formulated dG(0) method reduces to a time-stepping scheme which, in the present 'autonomous' case, is actually equivalent to the backward Euler scheme for the values $U_n := U_n^-$:

$$U_0 = u_0, \qquad U_n - U_{n-1} = k_n f(U_n), \quad n \geq 1.$$

Now, the a posteriori error analysis for the end-time error $\|e_N\|$ via duality argument proceeds as follows. We consider the (backward in time) dual problem

$$-z' - B(t)^* z = 0, \quad T \geq t \geq 0, \qquad z(T) = \|e_N\|^{-1} e_N, \tag{2.7}$$

with the operator

$$B(t) := \int_0^1 f_x'(U + se)\, ds,$$

or in weak formulation,

$$\sum_{n=1}^{N} \Big\{ \int_{I_n} (\varphi, -z' - B^* z)\, dt - (\varphi_n^-, [z]_n) \Big\} = 0. \tag{2.8}$$

Taking $\varphi := e$, and using integration by parts and Galerkin orthogonality, we conclude that, with an arbitrary $Z \in S_h^{(0)}$,

$$\begin{aligned}
\|e_N\| &= \sum_{n=1}^{N} \int_{I_n} (e, -z' - B^* z)\, dt - \sum_{n=1}^{N-1} (e_n^-, [z]_n) + (e_N^-, z_N^-) \\
&= \sum_{n=1}^{N} \int_{I_n} (e' - Be, z)\, dt + \sum_{n=2}^{N} ([e]_{n-1}, z_{n-1}^+) + (e_0^+, z_0^+) \\
&= \sum_{n=1}^{N} \Big\{ \int_{I_n} (e' - f(u) + f(U), z)\, dt + ([e]_{n-1}, z_{n-1}^+) \Big\} \\
&= \sum_{n=1}^{N} \Big\{ \int_{I_n} (f(U), z - Z)\, dt - ([U]_{n-1}, (z - Z)_{n-1}^+) \Big\}.
\end{aligned}$$

For the special choice $Z = \bar{z}$, the interval-wise mean value of z,

$$\bar{z}_{|I_n} := k_n^{-1} \int_{I_n} z \, dt,$$

we obtain the error representation

$$\|e_N\| = -\sum_{n=1}^{N} ([U]_{n-1}, (z-\bar{z})_{n-1}^{+}).$$

From this, we infer the following two types of a posteriori error estimates.

- Local 'weighted' a posteriori error estimate:

$$\|e_N\| \le c_I \sum_{n=1}^{N} \left\{ k_n \, \|k_n^{-1}[U]_{n-1}\| \int_{I_n} \|z'\| \, ds \right\} =: c_I \sum_{n=1}^{N} k_n \rho_n \omega_n, \qquad (2.9)$$

  with the 'interpolation constant' $c_I = \frac{1}{2}$. The weights $\omega_n = \int_{I_n} \|z'\| \, ds$ represent the sensitivity of the error $\|e_N\|$ with respect to the local residuals $\rho_n = \|k_n^{-1}[U]_{n-1}\|$.

- Global a posteriori error estimate:

$$\|e_N\| \le c_I \Big(\max_{1\le n\le N} \{k_n \rho_n\} \Big) \int_I \|z'\| \, dt =: c_I c_S \max_{1\le n\le N} \{k_n \rho_n\}. \qquad (2.10)$$

  The stability constant $c_S = \int_I \|z'\| \, dt$ represents the 'global' sensitivity of the error $\|e_N\|$ with respect to the maximal residual.

On the basis of the above estimates, we obtain the following 'implicit' a posteriori step-size control strategies (k_n the old and k_n' the new step size):

$$c_I N k_n \rho_n \omega_n \quad \text{or} \quad c_I c_S k_n \rho_n \left\{ \begin{array}{rcl} \ge \text{TOL} & \Rightarrow & k_n' = \frac{1}{2} k_n \\ \sim \text{TOL} & \Rightarrow & k_n' = k_n \\ \le \frac{1}{4}\,\text{TOL} & \Rightarrow & k_n' = 2k_n \end{array} \right\}.$$

Here, the weights ω_n and the stability constant c_S can be approximated by the following formula, given a numerical approximation Z of z:

$$\omega_n := \int_{I_n} \|z'\| \, dt \approx \|[Z]_n\|, \qquad c_S := \int_I \|z'\| \, dt \approx \sum_{n=1}^{N} \|[Z]_n\|.$$

Remark 2.2. The *a posteriori* error estimates (2.9) and (2.10) also have drawbacks which should be pointed out:

- The evaluation of the weights ω_n requires the solution of the dual problem over the whole time interval $[0, T]$.

- This 'global' step size control is 'implicit' since it involves the simultaneous adaptation of all time intervals $\{I_1, \ldots, I_N\}$ in each adaptation cycle.

Remark 2.3. Duality-based a posteriori error analysis for the Galerkin approximation of ODEs has also been considered in Estep and French [59], and Estep [58]. Here, the error estimates contain global stability constants which are derived by analytical arguments.

2.2 Efficiency comparison: FD versus FE method

In the following, we want to compare the efficiency of the step-size selection strategies described above for the FD and the FE method. We emphasize that in the present situation ('autonomous' ODE), both methods are only different ways of writing the very same scheme. Nevertheless, we will see that using either the FD or the FE approach gives quite different results. Consider the special scalar initial value problem

$$u'(t) = u(t)^2, \quad 0 \le t \le T < 1, \quad u(0) = 1,$$

with the singular solution (see Figure 2.1)

$$u(t) = \frac{1}{1-t}.$$

Let the goal of the computation again be the approximation of the end-time value $u(T)$. For the different mesh adaptation strategies designed for the backward Euler scheme and the dG(0) method, we want to estimate the work which is required for computing with accuracy TOL, in dependence on $T \to 1$.

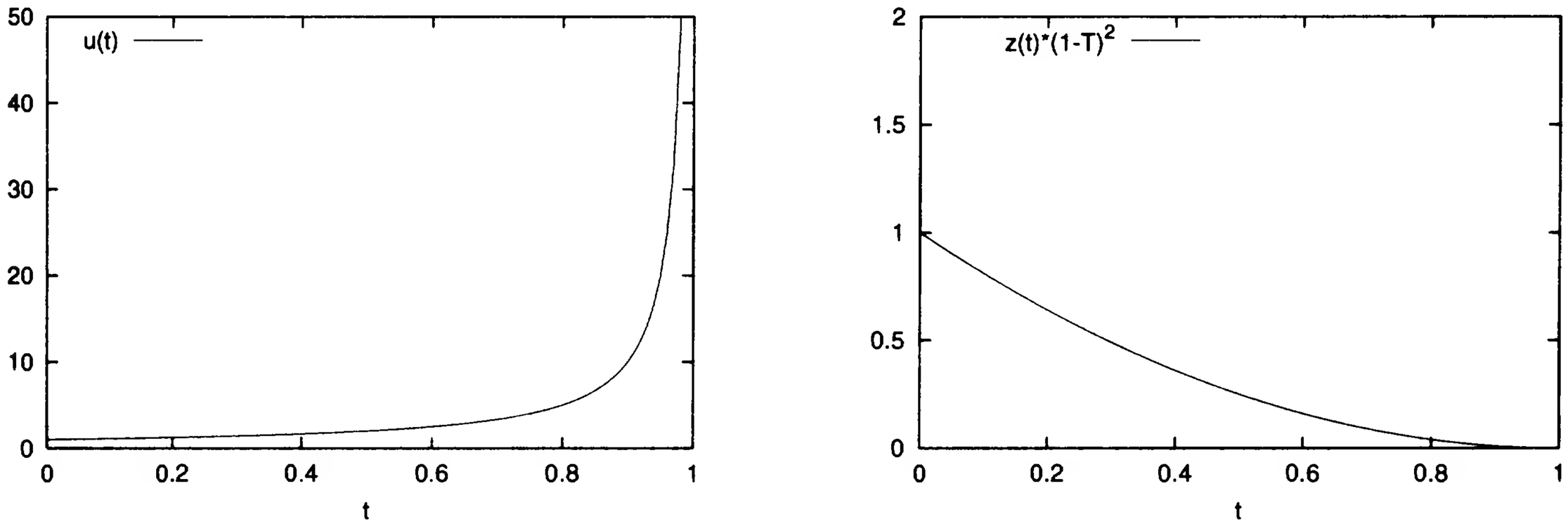


Figure 2.1: *Singular solution (left) and corresponding dual solution (right) for the evaluation of the end-time error.*

The dG(0) method

Let the step-size control be based on the *global* a posteriori error estimate (2.10),

$$|e_N| \le c_I c_S \max_{1\le n\le N}\{k_n\rho_n\}, \qquad \rho_n := k_n^{-1}|[U]_{n-1}|,$$

with the interpolation constant $c_I = \frac{1}{2}$ and the stability constant c_S from the dual a priori bound

$$\int_0^T |z'|\,dt =: c_S.$$

The criterion for the choice of the local step size is

$$c_I c_S k_n \rho_n = \mathrm{TOL} \quad \Leftrightarrow \quad k_n = \frac{\mathrm{TOL}}{c_I c_S \rho_n} \quad \Rightarrow \quad |e_N| \le \mathrm{TOL}.$$

Notice that, by construction, $\sum_{n=1}^N k_n = T$. For the explicit estimation of the work count, the parameters in this formula are approximated as follows:
i) Suppose that the step size is already small enough such that

$$\rho_n = k_n^{-1}|[U]_n| \approx k_n^{-1}|u_n - u_{n-1}| \approx \sup_{I_n}|u'| \approx (1-t_n)^{-2}.$$

ii) The stability constant c_S is essentially determined by the dual problem (after linearization about the continuous solution u)

$$z'(t) = -\frac{2}{1-t}z(t), \quad 1 > T \ge t \ge 0, \quad z(T) = 1,$$

with the solution

$$z(t) = \exp\left(2\int_t^T \frac{ds}{1-s}\right) z(T) = \left(\frac{1-t}{1-T}\right)^2.$$

Hence,

$$c_S = \int_0^T |z'|\,dt \le \frac{1}{(1-T)^2}.$$

This leads us to the step-size distribution

$$k_n \approx (1-t_n)^2(1-T)^2\,\mathrm{TOL}.$$

The corresponding work measure N is determined as

$$N = \sum_{n=1}^N k_n k_n^{-1} \approx \frac{1}{(1-T)^2\mathrm{TOL}}\sum_{n=1}^N k_n(1-t_n)^{-2} \approx \frac{1}{(1-T)^2\mathrm{TOL}}\int_0^T \frac{dt}{(1-T)^2},$$

and consequently

$$N \approx \frac{1}{(1-T)^3}\,\frac{1}{\mathrm{TOL}}. \tag{2.11}$$

For later purposes, we note that the local weights are determined by

$$\omega_n = \int_{I_n} |z'(t)|\,dt \;\approx\; k_n\,\frac{1-t_n}{(1-T)^2}.$$

The backward Euler scheme

Let the step-size control be based on the *global* a priori error estimate

$$|e_N| \;\le\; K\sum_{n=1}^{N} k_n|\tau_n(u)|, \qquad |\tau_n(u)| \le \tfrac{1}{2}k_n \sup_{I_n}|u''|.$$

The corresponding criterion for the control of the local step sizes is

$$k_n \;=\; \frac{\mathrm{TOL}}{KT\sup_{I_n}|u''|} \quad\Rightarrow\quad |e_N| \le \mathrm{TOL}.$$

The parameters in this formula are approximated as follows:
i) The Lipschitz constant of $f(\cdot)$ along the solution u is

$$L_f(t) = \frac{2}{1-t},$$

such that the growth factor becomes

$$K(T, L_f) \;\approx\; \exp\left(\int_0^T L_f(t)\,dt\right) \;\approx\; \exp\left(\int_0^T \frac{2}{1-t}\,dt\right) \;\approx\; (1-T)^{-2}.$$

ii) The leading truncation error $|\tau_n^0|$ behaves like

$$|\tau_n^0| \le \tfrac{1}{2}\sup_{I_n}|u''| \approx (1-t_n)^{-3}.$$

This results in the step-size distribution

$$k_n \;\approx\; \frac{\mathrm{TOL}}{TK(T)\sup_{I_n}|u''|} \;\approx\; (1-T)^2(1-t_n)^3\,\mathrm{TOL}.$$

The corresponding work count is

$$N = \sum_{n=1}^{N} k_n k_n^{-1} \;\approx\; \frac{1}{(1-T)^2}\,\frac{1}{\mathrm{TOL}}\sum_{n=1}^{N} k_n(1-t_n)^{-3},$$

and consequently (compare this to (2.11)),

$$N \approx \frac{1}{(1-T)^4}\,\frac{1}{\mathrm{TOL}}. \tag{2.12}$$

Remark 2.4. We note that for both FD-based and FE-based schemes the efficiency of the computation can be further improved by using a more localized adaptation strategy. This will be considered in the exercises below.

Remark 2.5. The critical drawback of the heuristic step-size control strategy used for the backward Euler scheme is the possible crude under-estimation of the growth factor $K(T, L_f)$. On the other hand, estimating $K(T, L_f)$ by an *a priori* analysis is oriented at the *worst case* scenario and usually leads to over-estimation rendering the resulting error bound useless. In general, this prevents the step-size control strategy based on the *a priori* error estimate (2.3) from providing a real *control* of the error. In order to overcome this limitation the following two different approaches may be used:
i) The first one is based on the relation

$$e_n = e_{n-1} + k_n f'(U_n)e_n + k_n\tau_n(u) + k_n O(e_n^2), \tag{2.13}$$

for the error $e_n = u_n - U_n$, with initial value $e_0 = 0$. Using a guess for the truncation error $\tau_n(U_n) \approx \tau_n(u)$, obtained for example by local extrapolation, the solution E_n of the *linearized* error equation

$$E_n = E_{n-1} + k_n f'(U_n)E_n + k_n\tau_n(U_n), \quad 0 \le n \le N, \tag{2.14}$$

may be used to get a guess for the true end-time error, $E_N \approx e_N$. However, this does not provide criteria for a time-step adaptation to reduce the error below the prescribed tolerance, $\|e_N\| \le TOL$.
ii) An alternative approach employs a duality argument similar to that one used for the dG(0) method, but now on the *discrete* level. Let Z_n be the solution of the linearized backward-in-time scheme

$$Z_{n-1} = Z_n + k_n f'(U_n)Z_{n-1}, \quad 0 \le t_n < t_N. \tag{2.15}$$

with starting value Z_N. Then, using the error relation (2.13), we obtain

$$\begin{aligned}(e_n, Z_n) &= (e_n, Z_n - Z_{n-1}) + (e_n - e_{n-1}, Z_{n-1}) + (e_{n-1}, Z_{n-1})\\ &= -k_n(e_n, f'(U_n)Z_{n-1}) + k_n(f'(U_n)e_n, Z_{n-1})\\ &\quad + k_n(\tau_n(u) + \mathcal{O}(e_n^2), Z_{n-1}) + (e_{n-1}, Z_{n-1}),\end{aligned}$$

and summing over $1 \le n \le N$,

$$(e_N, Z_N) = (e_0, Z_0) + \sum_{n=1}^{N} k_n(\tau_n(u) + \mathcal{O}(e_n^2), Z_{n-1}). \tag{2.16}$$

For $Z_N := e_N\|e_N\|^{-1}$ and $e_0 = 0$, we arrive at

$$\|e_N\| \le c_{S,k} \sum_{n=1}^{N} k_n\{\|\tau_n(u)\| + \mathcal{O}(\|e_n\|^2)\}, \tag{2.17}$$

with the *discrete* stability constant $c_{S,k} := \max_{0\le n\le N-1} \|Z_n\|$. Neglecting the quadratic error term, we obtain the error estimate

$$\|e_N\| \approx c_{S,k} \sum_{n=1}^{N} k_n \|\tau_n(U_n)\|, \tag{2.18}$$

with an approximation $\tau_n(U_n) \approx \tau_n(u)$ for the truncation error. Here, the generally too pessimistic *a priori* bound $K(t_N, L_f)$ is replaced by the *a posteriori* stability constant $c_{S,k}$. In practice, $c_{S,k}$ will be determined from the computed dual solution Z_n using again a suitable guess for the starting value $Z_N = e_N \|e_N\|^{-1}$. The step-size selection strategy based on the estimate (2.18), is generically *implicit* like that proposed for the dG(0) method, but is capable of producing useful error bounds. Surprisingly, it has not found much attention in the FD community.

2.3 Exercises

Exercise 2.1. Consider the autonomous initial value problem

$$u'(t) = u(t)^2, \quad 0 \le t \le T < 1, \quad u(0) = 1,$$

with the solution $u(t) = (1-t)^{-1}$. Compute $u(T)$ by the backward Euler scheme

$$U_n = U_{n-1} + k_n U_n^2, \quad n \ge 1, \quad U_0 = 1.$$

Let the step sizes k_n be chosen on the basis of the a priori error bound (2.3),

$$|e_N| \le K(T) \sum_{n=1}^{N} k_n \tau_n(u), \qquad \tau_n(u) = k_n \tau_n^0(u) + \mathcal{O}(k_n^2),$$

according to the *implicit* control

$$k_n \approx \left(\frac{\mathrm{TOL}}{N\, K(T) |\tau_n^0(u)|} \right)^{1/2}.$$

What is the asymptotic work count $N = N(\mathrm{TOL}, T)$ as $T \to 1$?

Exercise 2.2. Consider the model problem of Exercise 2.1. Let in the dG(0) method the step-size selection be based on the *weighted* error estimator (2.9),

$$|e_N| \le c_i \sum_{n=1}^{N} k_n \rho_n \omega_n,$$

according to the *implicit* control

$$k_n \approx \frac{\mathrm{TOL}}{N \rho_n \omega_n}.$$

What is the work count $N = N(\mathrm{TOL}, T)$ as $T \to 1$?

Exercise 2.3. The dG(0) method is only of first order. By the same principle, implicit dG(r) methods of any order $r \geq 1$ can be designed. Alternatives are the so-called *cG(r) (continuous Galerkin) methods* which are Petrov-Galerkin methods. The simplest representative is the second-order cG(1) method which uses (continuous) piecewise linear trial and (discontinuous) piecewise constant test functions:

$$U(0) = u_0, \quad \sum_{n=1}^{N} \int_{I_n} (U' - f(U), \psi)\, dt = 0 \qquad \forall \psi \in S_h^{(0)}(I).$$

This method is closely related to the implicit midpoint rule

$$U_0 = u_0, \quad U_n = U_{n-1} + k_n f(\tfrac{1}{2}\{U_{n-1} + U_n\}), \quad n \geq 1.$$

Develop a residual-based a posteriori error estimate for the end-time error $\|e_N\|$ of the cG(1) method.

Exercise 2.4 *(Practical exercise).* Verify the theoretical predictions in Exercises 2.1 and 2.2 by a computational experiment. Use the backward Euler scheme and the dG(0) method for approximating the solution value $u(T)$ in the model problem

$$u'(t) = u(t)^2, \quad 0 \leq t \leq T < 1, \quad u(0) = 1,$$

with the solution $u(t) = (1-t)^{-1}$.
a) Use all the step-size selection strategies developed in the text and in Exercises 2.1 and 2.2 with the formulas for $K(T)$, $|\tau_n^0(u)|$, ρ_n and ω_n as given in the text. Determine experimentally the number of resulting time steps N for a decaying sequence of tolerances $\mathrm{TOL_i} = 2^{-\mathrm{i}}$, $i = 1, 2, 3, \ldots$. Monitor the true error $e(T) = u(T) - U_N$ and compare it with the given tolerance. Interpret the observed results.
b) Approximate the leading truncation errors $|\tau_n^0|$ by second-order difference quotients of the computed solution U_n, and the residuals $\rho_n = k_n^{-1}|[U]_n|$, as defined in the text. Repeat the above test using these approximations and report the differences to the results observed in (a).
c) Do the same test as in (a) for a sequence of *uniform* step size distributions with $k \equiv T/N$. You will observe that the performance of equidistant time meshes is almost as good as that of the best adaptive procedure based on the 'weighted' a posteriori error estimate for the dG(0) method. This surprising phenomenon can be explained by showing that in this case

$$N \approx (1-T)^{-2} |\log(1-T)|\, \mathrm{TOL}^{-1}.$$

Try to detect this logarithmic behavior in the numerical experiment. So what is the benefit of sophisticated local time-step adaptation?
d) Perform the same experiment with the cG(1) method described in Exercise 2.3. First, use uniform step sizes and, then, try to use your a posteriori error indicator derived in Exercise 2.3. This should demonstrate the superiority of the second-order cG(1) over the only first-order dG(0) method.

Chapter 3

A PDE Model Case

In this chapter, we will develop the basics of the DWR method for linear elliptic partial differential equations as originally described in Becker and Rannacher [30]. As a model configuration, we consider the Poisson equation on a polygonal or polyhedral domain $\Omega \subset \mathbb{R}^d$, with Dirichlet boundary conditions:

$$-\Delta u = f \quad \text{in } \Omega, \quad u_{|\partial\Omega} = 0. \tag{3.1}$$

The discretization will be by a Galerkin finite element method which is based on the variational formulation of (3.1): Find $u \in V := H_0^1(\Omega)$, such that

$$a(u,\psi) := (\nabla u, \nabla\psi) = (f,\psi) \quad \forall \psi \in V. \tag{3.2}$$

In this chapter, we will exemplarily consider the a posteriori control of the resulting approximation u_h with respect to the following goals:

- Computation of an overview of the solution's structure (global norm error):

$$\|\nabla(u-u_h)\| \leq TOL, \qquad \|u-u_h\| \leq TOL.$$

- Computation of 'displacement' or 'stress' components at some point $a \in \bar{\Omega}$ (point-value error):

$$J(u) := u(a), \qquad J(u) := \partial_i u(a).$$

- Computation of 'mean normal flux':

$$J(u) := \int_{\partial\Omega} \partial_n u \, ds.$$

However, we will always have in mind the accurate computation with respect to *arbitrary* functional values $J(u)$ of the solution.

Here and below, we use the following notation: For a domain $\Omega \subset \mathbb{R}^d$, $L^2(\Omega)$ is the Lebesgue space of square-integrable functions on Ω, which is a Hilbert space with the scalar product and norm

$$(v,w)_\Omega = \int_\Omega v\, w\, dx, \qquad \|v\|_\Omega = \Big(\int_\Omega |v|^2\, dx\Big)^{1/2}.$$

Analogously, $L^2(\partial\Omega)$ is the space of square-integrable functions defined on the boundary $\partial\Omega$ equipped with the scalar product and norm

$$(v,w)_{\partial\Omega} = \int_{\partial\Omega} v\, w\, \mathrm{ds}, \qquad \|\mathrm{v}\|_{\partial\Omega} = \Big(\int_{\partial\Omega} |\mathrm{v}|^2\, \mathrm{ds}\Big)^{1/2}.$$

The Sobolev spaces $H^1(\Omega)$ and $H^2(\Omega)$ consist of those functions $v \in L^2(\Omega)$ which possess first- and second-order (distributional) derivatives $\nabla v \in L^2(\Omega)^d$ and $\nabla^2 v \in L^2(\Omega)^{d\times d}$, respectively. For functions in these spaces, we use the semi-norms

$$|v|_{1;\Omega} := \|\nabla v\|_\Omega, \qquad |v|_{2;\Omega} := \|\nabla^2 v\|_\Omega.$$

This notation can be extended to Sobolev spaces $H^p(\Omega)$ of arbitrary order $p \geq 1$. The space $H^1(\Omega)$ can be embedded in the space $L^2(\partial\Omega)$, such that for each $v \in H^1(\Omega)$ there exists a trace $v_{|\partial\Omega} \in L^2(\partial\Omega)$. Further, the functions in the subspace $H^1_0(\Omega) \subset H^1(\Omega)$ are characterized by the property $v_{|\partial\Omega} = 0$. By the Poincaré inequality,

$$\|v\|_\Omega \leq c\,\|\nabla v\|_\Omega, \quad v \in H^1_0(\Omega), \tag{3.3}$$

the H^1-semi-norm $\|\nabla v\|_\Omega$ is a norm on the subspace $H^1_0(\Omega)$. If the set Ω is identical with the domain on which the differential equation is posed, we usually omit the subscript Ω in the notation of norms and scalar products, for instance $\|v\| = \|v\|_\Omega$. All the above notation will be synonymously used also for vector- or matrix-valued functions $v : \Omega \to \mathbb{R}^d$ or $\mathbb{R}^{d\times d}$.

3.1 Finite element approximation

The discretization of the model problem (3.1) seeks an approximations $u_h \in V_h$, the so-called *Ritz projection* of u, in a certain finite dimensional subspace $V_h \subset V$,

$$a(u_h,\psi_h) = (f,\psi_h) \quad \forall \psi_h \in V_h. \tag{3.4}$$

The main feature of the Galerkin method for linear problems is the so-called *Galerkin orthogonality* for the error $e := u - u_h$,

$$a(e,\psi_h) = 0, \quad \psi_h \in V_h. \tag{3.5}$$

The subspaces (finite element spaces) considered have the form

$$V_h = \{v \in V : v_{|K} \in P(K),\ K \in \mathbb{T}_h\},$$

defined on decompositions $\mathbb{T}_h$ of Ω into *cells* K (triangles or quadrilaterals in $\mathbb{R}^2$, and tetrahedra or hexahedra in $\mathbb{R}^3$) of width $h_K = \mathrm{diam}(K)$; we write $h = \max_{K\in\mathbb{T}_h} h_K$ for the *global* mesh width. Here, $P(K)$ denotes a suitable space of polynomial-like functions defined on the cell $K \in \mathbb{T}_h$. In the numerical results discussed below, we have mostly used 'bilinear' or 'trilinear' finite elements on quadrilateral or hexahedral meshes, respectively, in which case $P(K) = \tilde{Q}_1(K)$ consists of shape functions obtained via a bilinear transformation from the space of 'bilinears' $Q_1(K) = \mathrm{span}\{1, x_1, x_2, x_1x_2\}$ or 'trilinears' $Q_1(K) = \mathrm{span}\{1, x_1, x_2, x_3, x_1x_2, x_2x_3, x_3x_1, x_1x_2x_3\}$ on the reference cell $\hat{K} = [0,1]^d$. Local mesh refinement or coarsening is realized by using *hanging nodes* in such a way that global conformity is preserved, that is $V_h \subset V$ (see also Section 4.2). For technical details of finite element spaces, the reader may consult the standard literature, for instance Ciarlet [46], Johnson [83] or Brenner and Scott [43], and especially Carey and Oden [44] for the treatment of hanging nodes.

We consider the control of the error with respect to some 'output functional' $J(\cdot)$, i.e., we want to have estimates for the difference $J(e) = J(u) - J(u_h)$. For simplicity, we assume here $J(\cdot)$ to be linear. Following the general concept of the DWR method, let $z \in V$ be the solution of the associated *dual problem*

$$a(\varphi, z) = J(\varphi) \quad \forall \varphi \in V, \tag{3.6}$$

and $z_h \in V_h$ its finite element approximations defined by

$$a(\varphi_h, z_h) = J(\varphi_h) \quad \forall \varphi_h \in V_h. \tag{3.7}$$

Using this construction together with Galerkin orthogonality, we obtain

$$\begin{aligned} J(e) &= a(e, z) = a(e, z-\psi_h) \\ &= (f, z-\psi_h) - a(u_h, z-\psi_h) =: \rho(u_h)(z-\psi_h), \quad \psi_h \in V_h. \end{aligned}$$

The so-called *residual* $\rho(u_h)(\cdot)$ of the Galerkin approximation u_h may be viewed as a functional on the solution space V. Cell-wise integration by parts implies

$$\begin{aligned} \rho(u_h)(z-\psi_h) &= \sum_{K\in\mathbb{T}_h} \{(f+\Delta u_h, z-\psi_h)_K - (\partial_n u_h, z-\psi_h)_{\partial K}\} \\ &= \sum_{K\in\mathbb{T}_h} \{(f+\Delta u_h, z-\psi_h)_K + \tfrac{1}{2}([\partial_n u_h], z-\psi_h)_{\partial K\setminus\partial\Omega}\}, \end{aligned}$$

where $[\partial_n u_h]$ denotes the jump of $\partial_n u_h$ across the inter-element edges (in 2-D) or faces (in 3-D), i.e., for two neighboring cells $K, K' \in \mathbb{T}_h$ with common edge Γ and normal unit vector n pointing from K to K', we set

$$[\partial_n u_h] = [\nabla u_h \cdot n] := (\nabla u_{h|K'\cap\Gamma} - \nabla u_{h|K\cap\Gamma}) \cdot n.$$

Since $n = -n'$, this actually defines the jump of $\partial_n u_h$ across the edge Γ. For later use, we define the cell and edge residuals R_h and r_h, respectively, by

$$R_{h|K} := f + \Delta u_h,$$
$$r_{h|\Gamma} := \begin{cases} \frac{1}{2}[\partial_n u_h], & \text{if } \Gamma \subset \partial K \backslash \partial\Omega, \\ 0, & \text{if } \Gamma \subset \partial\Omega. \end{cases}$$

We collect the previous results in the following Proposition.

Proposition 3.1. *For the finite element approximation (3.4) of the Poisson problem, we have the a posteriori error representation*

$$J(e) = \sum_{K \in \mathbb{T}_h} \left\{ (R_h, z - \psi_h)_K + (r_h, z - \psi_h)_{\partial K} \right\}, \tag{3.8}$$

with an arbitrary $\psi_h \in V_h$, and as a consequence the a posteriori error estimate

$$|J(e)| \leq \eta_\omega := \sum_{K \in \mathbb{T}_h} \rho_K \, \omega_K, \tag{3.9}$$

where the cell residuals ('smoothness indicators') ρ_K and weights ('influence factors') ω_K are given by

$$\rho_K := \left(\|R_h\|_K^2 + h_K^{-1} \|r_h\|_{\partial K}^2 \right)^{1/2},$$
$$\omega_K := \left(\|z - \psi_h\|_K^2 + h_K \|z - \psi_h\|_{\partial K}^2 \right)^{1/2},$$

for an arbitrary $\psi_h \in V_h$.

Remark 3.2. The *dual solution* z has the features of a 'generalized' Green function $G(K, K')$, as it describes the dependence of the target error quantity $J(e)$, which may be concentrated at some cell K, on local properties of the data, i.e. in this case the residuals $\rho_{K'}$ on cells K'; see Figure 3.1.

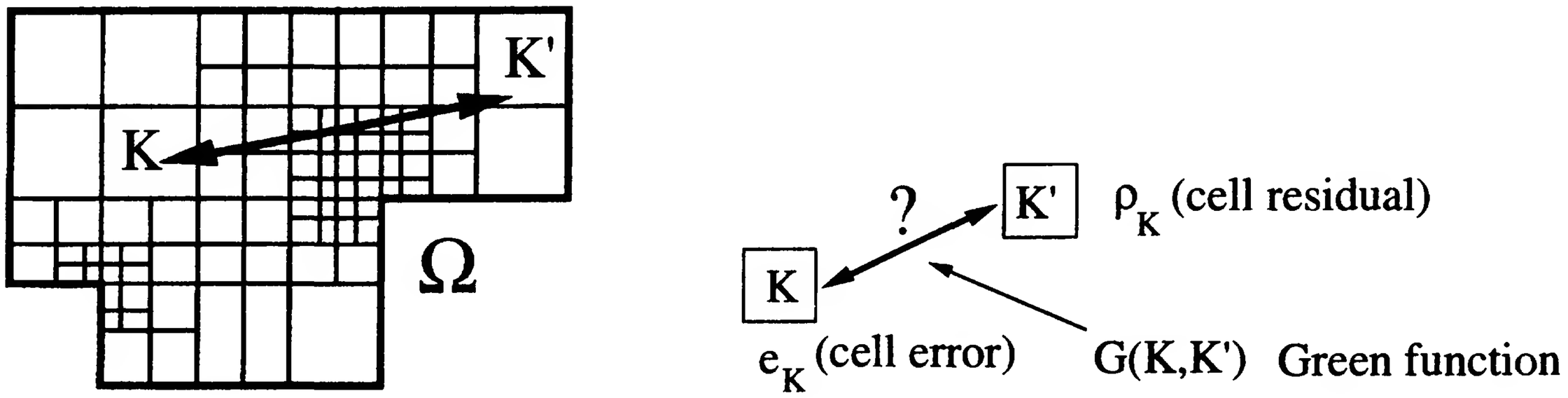


Figure 3.1: *Finite element mesh and scheme of error propagation.*

In order to evaluate the a posteriori error representation (3.8) or the resulting a posteriori error estimate (3.9), we need information about the 'continuous' dual solution z. Since in practice, z is not explicitly known, such information has to be obtained either through *a priori* analysis in form of bounds for z in certain Sobolev norms or through computation by solving the dual problem numerically. In the following, we will present some examples in which z can be bounded *a priori* or can even be explicitly determined. In later sections, dealing with real-life problems, we will always have to rely on the computational approximation of z.

3.2 Global a posteriori error estimates

In the following, we will demonstrate how the previous approach can be used for deriving the known a posteriori error estimates with respect to 'global' norms.

Example 3.1. (*Energy-norm error*) For deriving the usual error estimate with respect to the natural *energy norm* associated with problem (3.2), we choose the error functional

$$J(\varphi) = (\nabla\varphi, \nabla e)\|\nabla e\|^{-1},$$

considering the error e as a fixed quantity. The corresponding dual solution $z \in V$ is determined by

$$a(\varphi, z) = (\nabla\varphi, \nabla e)\|\nabla e\|^{-1} \quad \forall\varphi \in V,$$

and admits the trivial a priori bound $\|\nabla z\| \leq 1 =: c_S$ (stability constant). Clearly, in this particular case the dual solution is just given by $z = e\|\nabla e\|^{-1}$. From Proposition 3.1 we infer the estimate

$$J(e) = \|\nabla e\| \leq \sum_{K\in\mathbb{T}_h} \rho_K\,\omega_K \leq \Big(\sum_{K\in\mathbb{T}_h} h_K^2\rho_K^2\Big)^{1/2}\Big(\sum_{K\in\mathbb{T}_h} h_K^{-2}\omega_K^2\Big)^{1/2}.$$

We recall the interpolation error estimate (see, e.g., Scott and Zhang [125])

$$\inf_{\psi_h\in V_h}\Big(\sum_{K\in\mathbb{T}_h}\big\{h_K^{-2}\|z-\psi_h\|_K^2 + h_K^{-1}\|z-\psi_h\|_{\partial K}^2\big\}\Big)^{1/2} \leq \tilde{c}_I\,\|\nabla z\|, \tag{3.10}$$

to estimate further,

$$\|\nabla e\| \leq \tilde{c}_I\Big(\sum_{K\in\mathbb{T}_h} h_K^2\rho_K^2\Big)^{1/2}\|\nabla z\| \leq \tilde{c}_I\Big(\sum_{K\in\mathbb{T}_h} h_K^2\rho_K^2\Big)^{1/2}.$$

This results in the classical *energy-norm error estimate*

$$\|\nabla e\| \leq \eta_E := \tilde{c}_I\Big(\sum_{K\in\mathbb{T}_h} h_K^2\rho_K^2\Big)^{1/2}, \tag{3.11}$$

with ρ_K as defined in Proposition 3.1.

Remark 3.3. By Galerkin orthogonality, there holds

$$\begin{aligned}\|\nabla e\|^2 &= \|\nabla u\|^2 - 2(\nabla u, \nabla u_h) + \|\nabla u_h\|^2 \\ &= \|\nabla u\|^2 - 2(\nabla u_h, \nabla u_h) + \|\nabla u_h\|^2 = \|\nabla u\|^2 - \|\nabla u_h\|^2,\end{aligned}$$

such that *energy-norm error control* turns out to be equivalent to *energy error control*. However, this requires the energy form to be a scalar product.

Example 3.2. (*L^2-norm error*) To derive an estimate with respect to the L^2 norm, we choose the error functional

$$J(\varphi) = (\varphi, e)\|e\|^{-1}.$$

Suppose that the (polygonal or polyhedral) domain Ω is convex. Then, the corresponding dual solution $z \in V \cap H^2(\Omega)$ admits the a priori bound $\|\nabla^2 z\| \leq 1 =: c_S$ (stability constant). From the result of Proposition 3.1, we infer the estimate

$$\|e\| \leq \sum_{K\in\mathbb{T}_h} \rho_K \omega_K \leq \Big(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2\Big)^{1/2} \Big(\sum_{K\in\mathbb{T}_h} h_K^{-4} \omega_K^2\Big)^{1/2}.$$

Using the interpolation error estimate (see, e.g., Brenner and Scott [43])

$$\inf_{\psi_h\in V_h} \Big(\sum_{K\in\mathbb{T}_h} \{h_K^{-4}\|z-\psi_h\|_K^2 + h_K^{-3}\|z-\psi_h\|_{\partial K}^2\}\Big)^{1/2} \leq c_I \|\nabla^2 z\|, \tag{3.12}$$

we obtain

$$\|e\| \leq c_I \Big(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2\Big)^{1/2} \|\nabla^2 z\| \leq c_I c_S \Big(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2\Big)^{1/2}.$$

This results in the well-known *L^2-norm error estimate*

$$\|e\| \leq \eta_{L^2} := c_I c_S \Big(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2\Big)^{1/2}. \tag{3.13}$$

with ρ_K again as defined in Proposition 3.1. In comparison to the energy-norm error estimate (3.11) the L^2-norm estimate involves the weighting h_K^4 which reflects its higher order of convergence.

3.3 A posteriori error estimates for output functionals

Next, we turn to estimating the error with respect to local error functionals. Let TOL denote the accuracy we want to achieve.

Example 3.3. (*Point-value error*) To estimate the error at some point $a \in \Omega$, we use the regularized functional

$$J(u) := |B_\varepsilon|^{-1} \int_{B_\varepsilon} u \, dx = u(a) + \mathcal{O}(\varepsilon^2),$$

where B_ε is the ε-ball around the point a and $\varepsilon := TOL$. The corresponding dual solution z behaves like a regularized Green function, i.e. in 2-D:

$$z(x) = g_\varepsilon^a(x) \approx \log(r(x)), \quad r(x) := \sqrt{|x-a|^2+\varepsilon^2}.$$

By choosing ψ_h in the estimate (3.9) suitably, the weights have here the form

$$\omega_K \approx h_K^2 \|\nabla^2 z\|_K \approx h_K^2 |K|^{1/2} r_K^{-2}, \qquad r_K := \max_{x \in K} r(x),$$

such that

$$|e(a)| \approx \eta_\omega := c_I \sum_{K \in \mathbb{T}_h} \frac{h_K^3}{r_K^2} \rho_K. \tag{3.14}$$

Example 3.4. (*Point-value derivative error*) To estimate the error in the derivative in direction x_i at some interior point $a \in \Omega$, we use the regularized output functional

$$J(u) := |B_\varepsilon|^{-1} \int_{B_\varepsilon} \partial_i u \, dx = \partial_i u(a) + \mathcal{O}(\varepsilon^2),$$

where again $\varepsilon := TOL$. In this case the dual solution behaves like a regularized derivative Green function, i.e. in 2-D,

$$z(x) = \partial_i g_\varepsilon^a(x) \approx \frac{(x-a)_i}{r(x)^2}, \quad r(x) := \sqrt{|x-a|^2+\varepsilon^2},$$

and the corresponding weights like

$$\omega_K \approx h_K^2 \|\nabla^2 z\|_K \approx h_K^2 |K|^{1/2} r_K^{-3}.$$

This results in the a posteriori error estimate

$$|\partial_i e(a)| \approx \eta_\omega := c_I \sum_{K \in \mathbb{T}_h} \frac{h_K^3}{r_K^3} \rho_K. \tag{3.15}$$

Compared with (3.14), this localizes the region of influence towards the point a even more.

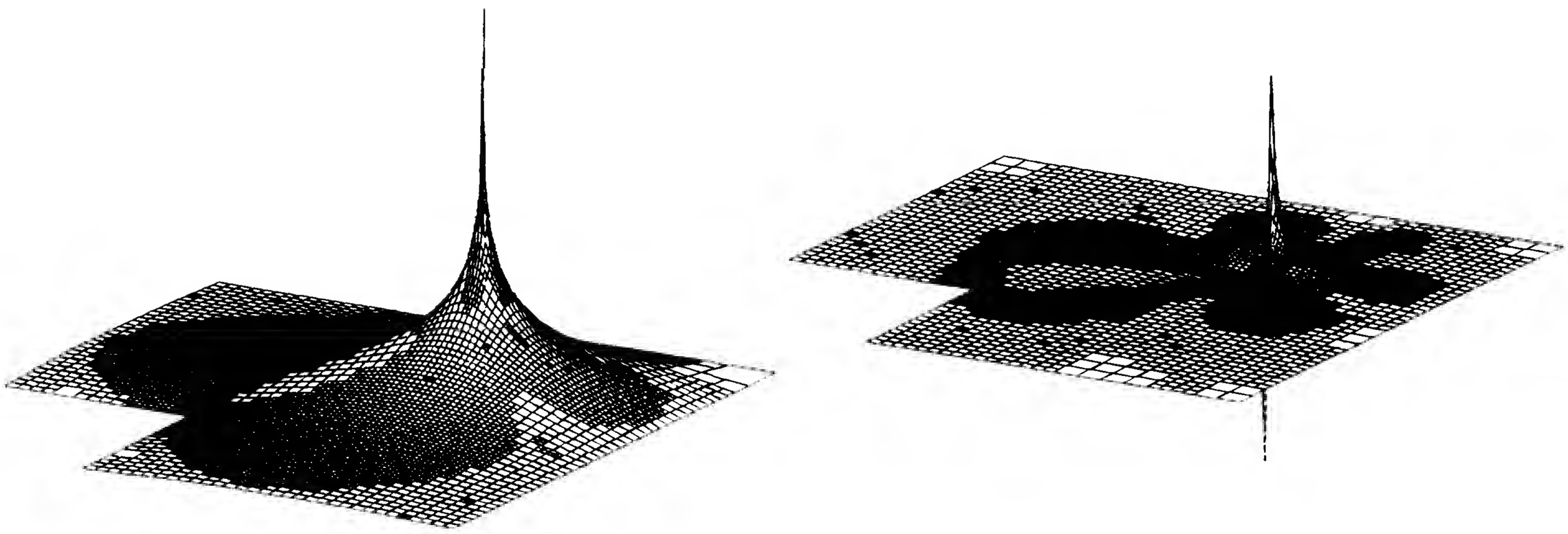


Figure 3.2: *Examples of computed dual solutions: regularized Green function and derivative Green function (scaled differently), for evaluating* $u(a)$ *and* $\partial_1 u(a)$ *.*

Example 3.5. (*Mean normal-flux error*) Another type of error functionals involves integrals over lower-dimensional manifolds. As an example, we consider the computation of the mean normal flux across the boundary,

$$J(u) = \int_{\partial\Omega} \partial_n u \, ds,$$

where for simplicity Ω is assumed to be the unit circle. The question is: *What is an 'efficient' mesh-size distribution for computing* $J(u)$*?* Notice that in this simple context the computational goal is trivial since it can be reduced to evaluating data, $J(u) = \int_\Omega \Delta u \, dx = -\int_\Omega f \, dx$. However, in more complex situations error functionals of this type cannot be so easily computed and are of high interest (c.f. the drag coefficient or the average Nusselt number mentioned in the Introduction). Here, the corresponding dual problem

$$a(\varphi, z) = (1, \partial_n \varphi)_{\partial\Omega} \quad \forall \varphi \in V \cap C^1(\bar{\Omega})$$

has a measure solution of the type $z \equiv -1$ in Ω, $z = 0$ on $\partial\Omega$. Hence, to avoid dealing with measures, we use the regularized output functional

$$J_\varepsilon(\varphi) = \varepsilon^{-1} \int_{S_\varepsilon} \partial_n \varphi \, dx = \int_{\partial\Omega} \partial_n \varphi \, ds + \mathcal{O}(\varepsilon),$$

where $S_\varepsilon = \{x \in \Omega : \ \mathrm{dist}\{x, \partial\Omega\} < \varepsilon\}$ and $\varepsilon := TOL$. The corresponding dual solution is explicitly given by

$$z_\varepsilon = \begin{cases} -1 & \text{in } \Omega \backslash S_\varepsilon, \\ -\varepsilon^{-1} \mathrm{dist}\{x, \partial\Omega\} & \text{in } S_\varepsilon. \end{cases}$$

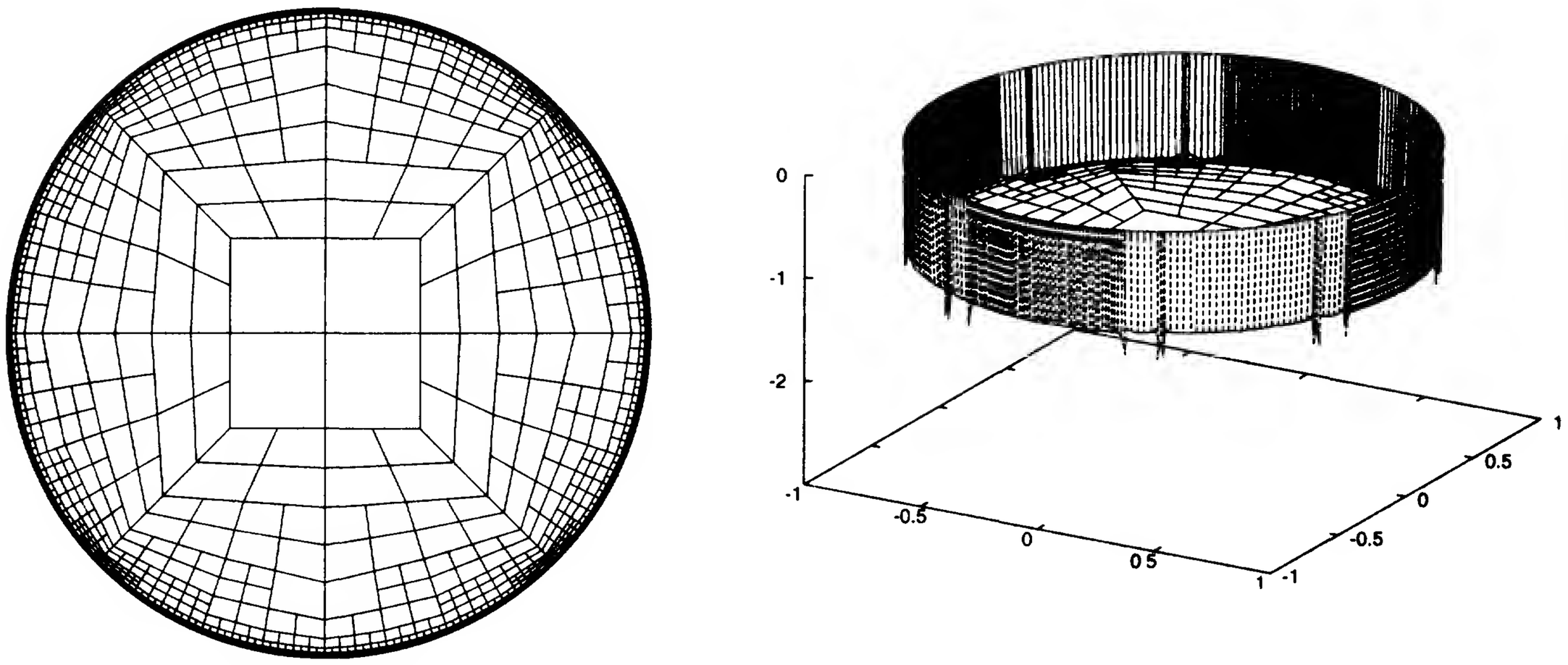


Figure 3.3: *Refined mesh and computed dual solution for the mean normal flux.*

On cells $K \subset \Omega \backslash S_\varepsilon$ there holds $z - I_h z \equiv 0$, which leads us to the error estimate

$$J_\varepsilon(e) \leq \eta_\omega := \sum_{K \in \mathbb{T}_h, K \cap S_\varepsilon \neq \emptyset} \rho_K \omega_K .$$

The conclusion is: *There is no contribution to the error from cells in the interior of* Ω. Hence, whatever right-hand side f, the optimal strategy is to refine the elements adjacent to the boundary and to leave the others unchanged. In practice, due to hanging nodes, this may also lead to some refinement in the interior, however.

Example 3.6. (*L^2-norm error*) Finally, we consider again the L^2-norm error estimate but for a problem with strongly varying diffusion coefficient,

$$-\nabla \cdot \{a \nabla u\} = f \quad \text{in } \Omega, \quad u_{|\partial\Omega} = 0. \tag{3.16}$$

This example is intended to show that even when estimating the error in global norms, it can be beneficial to keep the dual weights within the error estimator rather than condensing them into just one global stability constant. In the considered situation the dual problem reads in strong form

$$-\nabla \cdot \{a \nabla z\} = e \|e\|^{-1} \quad \text{in } \Omega, \quad z_{|\partial\Omega} = 0, \tag{3.17}$$

and the local residuals take the form

$$R_{h|K} = f + \nabla \cdot \{a \nabla u_h\}, \quad r_{h|\Gamma} = \tfrac{1}{2} n \cdot [a \nabla u_h].$$

By the same argument which led us to Proposition 3.1 and then to the standard L^2-error estimate, we infer the following two types of a posteriori error estimates:

- 'Weighted' error estimate:

$$\|e\| \leq \eta_{L^2}^{\omega} := c_I \sum_{K \in \mathbb{T}_h} h_K^2 \rho_K \omega_K, \quad \omega_K := \|\nabla^2 z\|_K. \tag{3.18}$$

- 'Global' error estimate:

$$\|e\| \leq \eta_{L^2} := c_I c_S \Big(\sum_{K \in \mathbb{T}_h} h_K^4 \, \rho_K^2 \Big)^{1/2}, \quad c_S := \|\nabla^2 z\|_\Omega. \tag{3.19}$$

The residual terms ρ_K are defined as before. In both cases, the stability terms can be evaluated by replacing the true dual solution by its Galerkin approximation $\|\nabla^2 z\|_K \approx \|\nabla_h^2 z_h\|_K$, with some second-order difference operator ∇_h^2. The interpolation constant is typically of size $c_I \approx 0.2$. The error-dependent functional $J(\cdot) = (\cdot, e)\|e\|^{-1}$ is evaluated by replacing the unknown solution u by a patch-wise higher-order interpolation $I_{2h}^{(2)} u_h$ of u_h,

$$e \approx I_{2h}^{(2)} u_h - u_h.$$

This gives us approximate L^2-error estimators denoted by $\tilde{\eta}_{L^2}^{\omega}(u_h)$ and $\tilde{\eta}_{L^2}(u_h)$, respectively. We want to compare the performance of these two L^2-error estimators by a numerical experiment. To this end, consider the particular setting $\Omega = (-1,1)^2$ and $a(x) = 0.1 + e^{3(x_1+x_2)}$, with a sinusoidal solution $u(x)$ and corresponding right-hand side f. In this calculation the mesh adaptation tries to equilibrate the local 'error indicators' $\eta_K = h_K^2 \rho_K \omega_K$ and $\eta_K = h_K^4 \rho_K^2$, respectively. (This and alternative strategies for mesh adaptation will be discussed in more detail in the next chapter.)

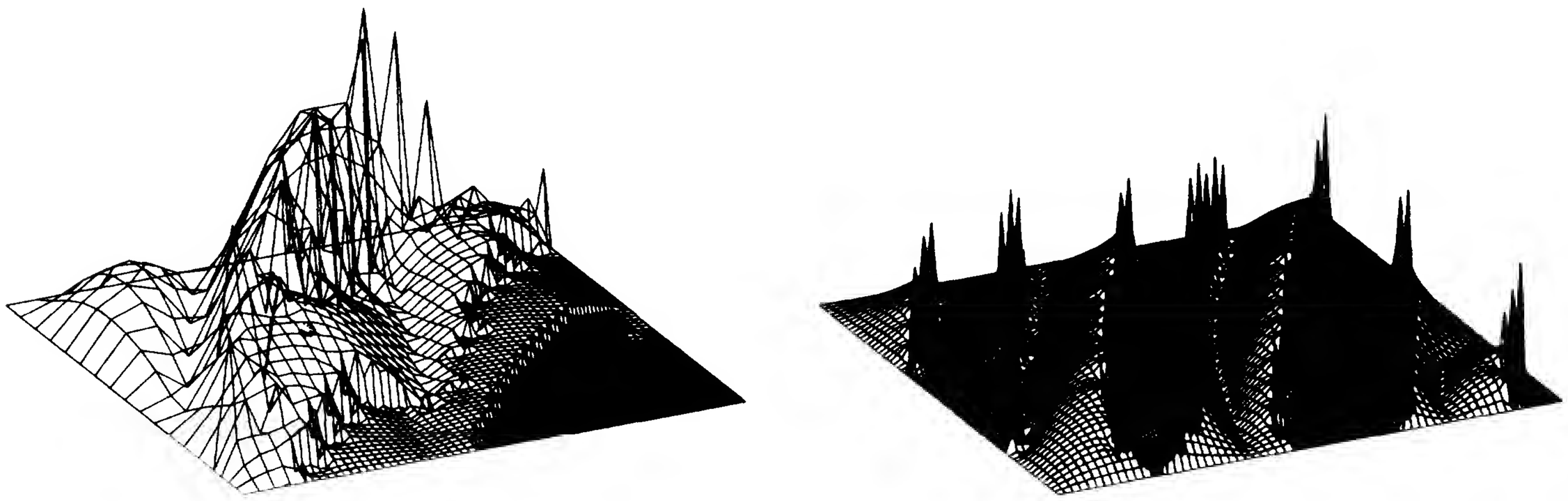


Figure 3.4: *Point-value errors obtained by η_{L^2} (left) and $\eta_{L^2}^{\omega}$ (right, scaled by 1:3) on meshes with $N \approx 10,000$ cells; from Becker and Rannacher [30].*

Table 3.1: *Results obtained by $\tilde{\eta}_{L_2}$ and $\tilde{\eta}^{\omega}_{L_2}$; from Becker and Rannacher [30].*

	$\tilde{\eta}_{L^2}$				$\tilde{\eta}^{\omega}_{L^2}$		
TOL	N	$\|e\|$	$\tilde{\eta}_{L^2}$	$\tilde{c}_s$	N	$\|e\|$	$\tilde{\eta}^{\omega}_{L^2}$
4^{-2}	2836	$6.40{\cdot}10^{-2}$	$2.32{\cdot}10^{-1}$	3.62	64	$1.47{\cdot}10^{-1}$	$1.52{\cdot}10^{-1}$
4^{-3}	5884	$2.13{\cdot}10^{-2}$	$1.21{\cdot}10^{-1}$	5.68	148	$1.08{\cdot}10^{-1}$	$9.80{\cdot}10^{-2}$
4^{-4}	15736	$7.36{\cdot}10^{-3}$	$4.76{\cdot}10^{-2}$	6.46	220	$6.77{\cdot}10^{-2}$	$5.24{\cdot}10^{-2}$
4^{-5}	23380	$5.59{\cdot}10^{-3}$	$3.12{\cdot}10^{-2}$	5.58	592	$2.21{\cdot}10^{-2}$	$2.59{\cdot}10^{-2}$
4^{-6}					892	$1.19{\cdot}10^{-2}$	$1.54{\cdot}10^{-2}$
4^{-7}					2368	$5.11{\cdot}10^{-3}$	$7.17{\cdot}10^{-3}$
4^{-8}					3640	$2.53{\cdot}10^{-3}$	$3.72{\cdot}10^{-3}$

Figure 3.4 shows the (scaled) error distribution on meshes obtained by the two estimators. The results shown in Table 3.1 indicate that efficient control of the L^2-norm error in the case of heterogeneous coefficients requires the use of 'weighted' a posteriori error estimates, i.e., the dual weights should be explicitly kept in the estimator and evaluated computationally.

Remark 3.4. (*Curved boundaries*) All examples considered so far had been posed on polygonal domains which can be exactly matched by the finite element mesh domain, i.e.,

$$\Omega_h := \cup\{K \in \mathbb{T}_h\} = \bar{\Omega}.$$

This assumption largely simplifies the error analysis and the resulting error estimators. However, in many practical cases at least parts of the boundary of the domain are curved and cannot be matched exactly by a polynomial approximation, e.g., in the cylinder-flow examples presented in the Introduction. Therefore, we have to deal with this complication.

Figure 3.5: *Standard situations of cells with curved edges.*

We consider the two typical situations depicted by Figure 3.5 in which a cell K at the boundary has a curved edge $\Gamma_K \subset \partial\Omega$. The computational domain can be made to satisfy $\Omega_h = \bar{\Omega}$, by simply extending ot truncating the domain of definition of shape functions. Shape functions and transformations are left unchanged. This approximation results in a *non-conforming* finite element scheme,

$$(\nabla u_h, \nabla \psi_h) = (f, \psi_h) \quad \forall \psi_h \in V_h, \tag{3.20}$$

in which $V_h \not\subset V$, since the elements of V_h will usually not satisfy zero boundary conditions. For the error analysis, we assume that the error functional $J(\cdot)$ has an L^2 representation, i.e., there is a $j \in L^2(\Omega)$, such that $J(\varphi) = (\varphi, j)$. Situations in which this is not the case require special considerations.

Using the solution $z \in V$ of the dual problem

$$-\Delta z = j \quad \text{in } \Omega, \quad z_{|\partial\Omega} = 0, \tag{3.21}$$

we obtain the error identity

$$J(e) = (j, e) = (e, -\Delta z) = (\nabla e, \nabla z) - (e, \partial_n z)_{\partial\Omega}.$$

For any $\psi_h \in V_h$, there holds

$$\begin{aligned}(\nabla e, \nabla \psi_h) &= (\nabla u, \nabla \psi_h) - (\nabla u_h, \nabla \psi_h) \\ &= (-\Delta u, \psi_h) + (\partial_n u, \psi_h)_{\partial\Omega} - (f, \psi_h) = (\partial_n u, \psi_h)_{\partial\Omega},\end{aligned}$$

and consequently,

$$J(e) = (\nabla e, \nabla(z - \psi_h)) - (\partial_n u, z - \psi_h)_{\partial\Omega} + (u_h, \partial_n z)_{\partial\Omega}.$$

Now, integrating by parts on each cell $K \in \mathbb{T}_h$, we obtain

$$\begin{aligned}J(e) = \sum_{K \in \mathbb{T}_h} &\left\{ (f + \Delta u_h, z - \psi_h)_K + (\partial_n e, z - \psi_h)_{\partial K} \right\} \\ &- (\partial_n u, z - \psi_h)_{\partial\Omega} + (u_h, \partial_n z)_{\partial\Omega}.\end{aligned}$$

This implies the error representation

$$J(e) = \sum_{K \in \mathbb{T}_h} \left\{ (R_h, z - \psi_h)_K + (r_h, z - \psi_h)_{\partial K} \right\} + (u_h, \partial_n z)_{\partial\Omega}, \tag{3.22}$$

with cell and edge residuals defined as above by $R_{h|K} := f + \Delta u_h$ and

$$r_{h|\Gamma} := \begin{cases} \frac{1}{2}[\partial_n u_h] & \text{if } \Gamma \subset \partial K \backslash \partial\Omega, \\ -\partial_n u_h & \text{if } \Gamma \subset \partial\Omega. \end{cases}$$

We note that the situation considered above is not very practical since the evaluation of the discrete equations (3.20) requires the use of numerical integration which in general introduces further errors that are not considered here.

Remark 3.5. (*Nonhomogeneous Dirichlet data*) Nonhomogeneous Dirichlet data

$$u = g \quad \text{on } \partial\Omega,$$

are usually treated by introducing a representative function $\hat{u} \in H^1(\Omega)$ satisfying $\hat{u}_{|\partial\Omega} = g$, and a corresponding finite element approximation $\hat{u}_h$ which is the interpolation (or the L^2 projection) of g along $\partial\Omega$. The solution is then sought as $u_h = \hat{u}_h + u_h^0$, where u_h^0 again has zero boundary values. Then, assuming again the domain to be polygonal or polyhedral, the error representation (3.8) takes the form

$$J(e) = \sum_{K \in \mathbb{T}_h} \left\{ (R_h, z - \psi_h)_K + (r_h, z - \psi_h)_{\partial K} \right\} - (g - g_h, \partial_n z)_{\partial\Omega}. \tag{3.23}$$

Remark 3.6. (*Neumann boundary conditions*) The treatment of Neumann boundary conditions

$$\partial_n u = g \quad \text{on } \Gamma_N \subset \partial\Omega,$$

does not cause any problems even in the case of a curved boundary, provided that the mesh $\mathbb{T}_h$ is compatible with the decomposition of the boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$. In this case, we simply assume the cells adjacent to the Neumann boundary to have possibly a curved edge or face matching the boundary exactly. Now, the variational formulation reads

$$a(u, \psi) = (f, \psi) + (g, \psi)_{\partial\Omega_N} \quad \forall \psi \in V,$$

where $V := \{v \in H^1(\Omega),\, v_{|\Gamma_D} = 0\}$. For the error of the corresponding Galerkin approximation, again Galerkin orthogonality holds, i.e., $a(e, \psi_h) = 0$ for $\psi_h \in V_h \subset V$. Then, the error representation (3.8) remains valid with the only modification that the edge residuals are now defined by

$$r_{h|\Gamma} := \begin{cases} \frac{1}{2}[\partial_n u_h], & \text{if } \Gamma \subset \partial K \backslash \partial\Omega, \\ 0, & \text{if } \Gamma \subset \Gamma_D, \\ g - \partial_n u_h, & \text{if } \Gamma \subset \Gamma_N. \end{cases}$$

3.4 Higher-order finite elements

We briefly describe the use of the DWR method for higher-order finite elements and will see that it can be used in this case without essential changes. Let $V_h^{(p)} \subset V$, be finite element spaces of order $p+1$, i.e., they possess the local approximation properties of polynomials of degree p. We recall that by setting $\psi_h = I_h^{(p)} z$ in (3.8), we have:

$$J(e) = \sum_{K \in \mathbb{T}_h} \left\{ (R_h, z - I_h^{(p)} z)_K + (r_h, z - I_h^{(p)} z)_{\partial K} \right\},$$

with the cell- and edge-residuals R_h and r_h as defined above and some local interpolation $I_h^{(p)} z \in V_h^{(p)}$. In order to extract cell-error indicators for local mesh adaptation, we may proceed as before in the low-order case:

$$\begin{aligned} |J(e)| &\leq \sum_{K\in\mathbb{T}_h} \left|(R_h, z-I_h^{(p)}z)_K + (r_h, z-I_h^{(p)}z)_{\partial K}\right| \\ &\leq \sum_{K\in\mathbb{T}_h} \left\{\|R_h\|_K \|z-I_h^{(p)}z\|_K + \|r_h\|_{\partial K}\|z-I_h^{(p)}z\|_{\partial K}\right\}. \end{aligned}$$

This does not reduce the asymptotic efficiency, as will be be demonstrated for the special case $p=2$. First, we collect the following estimates for the cell residuals and weights which are obtained by using standard trace estimates:

$$\begin{aligned} \|R_h\|_K &= \|f+\Delta u_h\|_K = \|\Delta e\|_K \\ h_K^{-1/2}\|r_h\|_{\partial K} &= \tfrac{1}{2} h_K^{-1/2}\|[\partial_n e]\|_{\partial K} \\ &\leq c h_K^{-1/2}\{h_K^{-1/2}\|\nabla e\|_{\tilde K} + h_K^{1/2}\|\nabla^2 e\|'_{\tilde K}\} \\ &\leq c h_K^{-1}\|\nabla e\|_{\tilde K} + \|\nabla^2 e\|'_{\tilde K}, \end{aligned}$$

where the prime in $\|\cdot\|'_{\tilde K}$ refers to a cell-wise evaluation of the norm, and $\tilde K$ is a patch of cells around K. Further there holds the higher-order interpolation estimate

$$\|z-I_h^{(p)}z\|_K + h_K^{1/2}\|z-I_h^{(p)}z\|_{\partial K} \leq c h_K^k \|\nabla^k z\|_{\tilde K}, \quad k=1,2,3.$$

Using the above estimates, we obtain on a quasi-uniform mesh, i.e. for $h_K \approx h$, that

$$|J(e)| \leq c\{h^{-1}\|\nabla e\| + \|\nabla^2 e\|'\} h^k \|\nabla^k z\| \leq c h^{k+1}\|\nabla^3 u\|\,\|\nabla^k z\|, \quad k=1,2,3.$$

We evaluate this error estimate for the special output functionals

$$J(\varphi) := (\nabla\varphi, \nabla e)\|\nabla e\|^{-1}, \quad J(\varphi) := (\varphi, e)\|e\|^{-1},$$

and

$$J_\psi(\varphi) := (\varphi, \psi)\|\nabla\psi\|^{-1}, \quad \psi \in H_0^1(\Omega),$$

which correspond to the energy-norm error, the L^2-norm error, and the error in the H^{-1} norm $\|\cdot\|_{-1}$, i.e. the norm of the dual space $H^{-1}(\Omega)$ of $H_0^1(\Omega)$. In virtue of the corresponding a priori bounds for the dual solution (for sufficiently regular domains Ω), we obtain the L^2 and energy error estimate

$$\|e\| + h\|\nabla e\| \leq c h^3 \|\nabla^3 u\|, \tag{3.24}$$

as well as the 'negative-norm' error estimate

$$\|e\|_{-1} := \sup_{\psi\in H_0^1(\Omega)} \frac{(e,\psi)}{\|\nabla\psi\|} \leq c h^4 \|\nabla^3 u\|, \tag{3.25}$$

which are all of optimal order.

3.5 Exercises

Exercise 3.1. Functional-oriented a posteriori error estimates can also be stated in terms of energy-norm error bounds. Consider the Poisson model problem. With the primal and dual errors $e := u - u_h$ and $e^* := z - z_h$, respectively, there holds

$$|J(e)| = (\nabla e, \nabla z) = (\nabla e, \nabla e^*) \leq \|\nabla e\| \, \|\nabla e^*\|$$

Then, any a posteriori bound for the energy-norm error supplies also a bound for $J(e)$. Specify a situation in which this simple minded approach is inefficient. Why is this approach not suited to extract refinement indicators from the error estimate?

Exercise 3.2. Let $\Omega \subset \mathbb{R}^2$ be a convex polygonal domain. Develop a residual-based a posteriori estimate for the error $e := u - u_h$ with respect to the L^∞-norm employing a *global* stability constant. Use the weighted a priori estimate

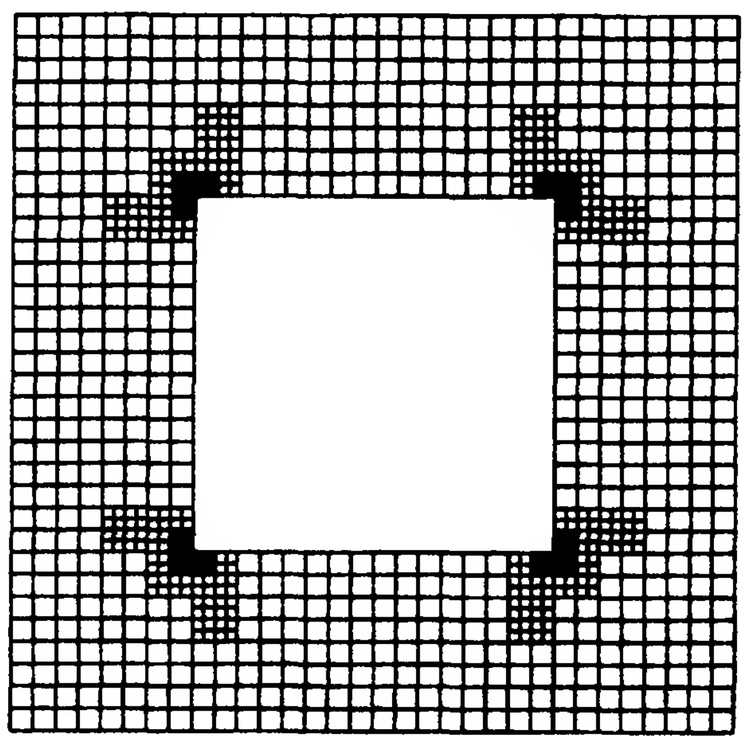
$$\|r \nabla^2 g_\varepsilon^a\| \leq c_\Omega \, |\log(\varepsilon)|,$$

for the regularized Green function g_ε^a.

Exercise 3.3 *(Practical exercise).* Consider the Poisson problem

$$-\Delta u = 1 \ \text{ in } \Omega, \quad u_{|\partial\Omega} = 0,$$

on the domain $\Omega \subset (-1,1)^2$, shown in the figure below. Compute the function values $u(a)$ and $\partial_1 u(a)$ at the point $a = (.75, .75)$.



a) Use the provided program for computing $\partial_1 u(a)$ for a sequence of tolerances $\mathrm{TOL}_i = 4^{-i}$, $i = 1, 2, \ldots$, and monitor the behavior of the true error and the number of cells depending on the achieved accuracy TOL. The code uses the duality-based weighted error estimator described in these Lecture Notes and the *fixed error fraction strategy* for mesh refinement.

b) Implement your own strategy of mesh refinement based on error indicators, smoothness indicators, or a priori information of your choice and try to beat the efficiency of the automatic error control process of (a).

Chapter 4

Practical Aspects

In this chapter, we discuss several aspects of the practical use of the DWR method described in the previous sections. These are (i) the practical and efficient evaluation of the a posteriori error representations, (ii) the extraction of local refinement indicators, and (iii) the design of strategies for economical mesh adaptation.

The starting point is the a posteriori *error representation*

$$J(e) = \sum_{K \in \mathbb{T}_h} \left\{ (R_h, z - \psi_h)_K + (r_h, z - \psi_h)_{\partial K} \right\} =: E(u_h). \tag{4.1}$$

as derived above for the Poisson problem. For its evaluation, due to Galerkin orthogonality, the subtraction of the arbitrary element $\psi_h \in V_h$ could be suppressed. From this, we extract local 'refinement indicators' called η_K of the form

$$\eta_K := \left| (R_h, z - \psi_h)_K + (r_h, z - \psi_h)_{\partial K} \right|,$$

which are used to steer the mesh adaptation. Collecting these error indicators, we obtain an a posteriori *error estimator*, i.e. an upper bound for the error,

$$|J(e)| \leq \eta := \sum_{K \in \mathbb{T}_h} \eta_K. \tag{4.2}$$

We emphasize that in practice it does not make much sense to estimate further in the indicators η_K, since we would inevitably lose sharpness of the error bound.

For practical use of the error representation (4.1) and the error bound (4.2), we have to approximate the terms involving the unknown dual solution z resulting in approximate error representations $\tilde{E}(u_h)$ and error indicators $\tilde{\eta}_K$. This may be expensive while the evaluation of the residuals R_h and r_h is usually cheap. We have to distinguish two related questions:

- Sharpness of the approximate error representations $\tilde{E}(u_h)$?
- Effectivity of the approximate local error indicators $\tilde{\eta}_K$ for mesh refinement?

For measuring the accuracy of the resulting error estimators, we will use the so-called *(reciprocal) effectivity index* defined by

$$I_{\text{eff}} := \left| \frac{\tilde{E}(u_h)}{J(e)} \right|,$$

which represents the degree of overestimation and should desirably be close to one.

Remark 4.1. Often the physical quantity to be computed can be expressed in different forms which coincide on the 'continuous' level but differ from each other for 'discrete' functions and may lead to more or less robust approximations. A typical example is the mean normal flux, and, more interesting, the drag and lift coefficients computed from solutions of the Navier-Stokes equations. Both can be expressed as surface or, after integration by parts, as volume integrals. If the desired output functional is not properly defined on the solution space V, such as point evaluation in two or more dimensions, it requires regularization,

$$J_\varepsilon(u) = J(u) + o(\varepsilon), \quad \varepsilon = TOL.$$

In the following, we will mostly suppress the index ε indicating this regularization.

Remark 4.2. In general, the dual problem has to be solved numerically. In the case of a nonlinear problem the first step is linearization which will be discussed in Chapter 6. Then, this 'linearized' dual problem is approximately solved in some finite element space $\tilde{V}_h \subset V$,

$$\tilde{z}_h \in \tilde{V}_h : \quad a'(u_h)(\varphi_h, \tilde{z}_h) = J(\varphi_h) \quad \forall \varphi_h \in \tilde{V}_h,$$

where not necessarily $\tilde{V}_h = V_h$. This will be discussed in more detail below.

Remark 4.3. If on the basis of a numerical approximation to the dual solution z an approximate error representation $\tilde{E}(u_h)$ has been generated, one may hope to obtain an improved approximation to the target quantity by setting

$$\tilde{J}(u_h) := J(u_h) + \tilde{E}(u_h) \approx J(u).$$

This 'post-processing' can significantly improve the accuracy in computing $J(u)$, but the resulting error can then no more be estimated on the basis of the available information.

4.1 Evaluation of the error identity and indicators

Consider the approximation of the model Poisson problem

$$-\Delta u = f \ \text{ in } \Omega, \quad u_{|\partial\Omega} = 0, \tag{4.3}$$

on a polygonal (or polyhedral) domain $\Omega \subset \mathbb{R}^d$ $(d = 2 \text{ or } 3)$ by piecewise bilinear (or trilinear) finite elements (in short 'Q_1 elements'). For this prototypical case,

we will compare several common strategies of evaluating the error representation (4.1) and the corresponding local error indicators. Notice that the approximation of z in $E(u_h)$ simply by its Ritz projection $z_h \in V_h$ does not work since, by Galerkin orthogonality, it would result in the useless approximation $\tilde{E}(u_h) = 0$.

Approximation by a higher-order method

A first possibility is to solve dual problem by using *biquadratic* finite elements on the current mesh yielding an approximation $z_h^{(2)} \in V_h^{(2)}$ to z. This yields the approximate error representation

$$E^{(1)}(u_h) := \sum_{K \in \mathbb{T}_h} \left\{ \left(R_h, z_h^{(2)} - I_h z_h^{(2)}\right)_K + \left(r_h, z_h^{(2)} - I_h z_h^{(2)}\right)_{\partial K} \right\},$$

and the corresponding local error indicators

$$\eta_K^{(1)} = \left| \left(R_h, z_h^{(2)} - I_h z_h^{(2)}\right)_K + \left(r_h, z_h^{(2)} - I_h z_h^{(2)}\right)_{\partial K} \right|.$$

For the special situation considered below (see Table 4.1), this approximation results in an asymptotically optimal effectivity index, $\lim_{TOL \to 0} I_{\text{eff}}^{(1)} = 1$. This observed behavior is supported by the theoretical analysis presented in the next section. However, approximating the dual problem by a higher-order method than used for u_h seems not very attractive and can actually be avoided in most cases. A compromise would be to compute only the residual with respect to the coarse-grid space $V_{2h}^{(2)}$ (biquadratics on mesh $\mathbb{T}_{2h}$) of the bilinear Ritz projection $z_h \in V_h$ and to perform a few steps of defect correction using the system matrix of V_h with appropriate blockwise preconditioning.

Approximation by higher-order interpolation

A simplification is achieved by patch-wise higher-order interpolation of the bilinear Ritz projection $z_h \in V_h$ of z. Here, we only consider *biquadratic* interpolation in 2-D. On square blocks of four neighboring cells the 9 nodal values of z_h are used to define a biquadratic interpolation $I_{2h}^{(2)} z_h$. This is then used in the error representation instead of z, resulting in the approximate error representation

$$E^{(2)}(u_h) := \sum_{K \in \mathbb{T}_h} \left\{ \left(R_h, I_{2h}^{(2)} z_h - z_h\right)_K + \left(r_h, I_{2h}^{(2)} z_h - z_h\right)_{\partial K} \right\},$$

and the corresponding local error indicators

$$\eta_K^{(2)} = \left| \left(R_h, I_{2h}^{(2)} z_h - z_h\right)_K + \left(r_h, I_{2h}^{(2)} z_h - z_h\right)_{\partial K} \right|.$$

For the special situation considered below (see Table 4.1), this approximation also results in an almost optimal effectivity index, $\underline{\lim}_{TOL \to 0} I_{\text{eff}}^{(2)} \sim 1$.

Remark 4.4. In the case of higher-order finite elements , with $p \geq 2$, the evaluation of the error identity (4.1) may be done in a similar way as for $p = 1$ employing a patch-wise interpolation $I_{2h}^{(p')} z_h$, with $p' > p$, of the Ritz projection $z_h \in V_h^{(p)}$. For example, in the case of biquadratic elements ($p = 2$), on a 2×2-cell patch we have 25 nodal values which can be used to construct an interpolation of degree $p = 4$, i.e., in this case we would use $I_{2h}^{(p+2)} z_h$.

Approximation by difference quotients

The error representation (4.1) is estimated by

$$|E(u_h)| \leq \sum_{K \in \mathbb{T}_h} \rho_K \, \omega_K,$$

with the notation of Proposition 3.1. Applying the usual cell-wise interpolation estimates, we have

$$\omega_K^2 = \|z - I_h z\|_K^2 + h_K \|z - I_h z\|_{\partial K}^2 \leq c_I^2 h_K^4 \|\nabla^2 z\|_K^2,$$

with an interpolation constant $c_I \sim 0.1 \ldots 1$. Now, the second derivatives $\nabla^2 z$ are replaced by suitable second-order difference quotients $\nabla_h^2 z_h$ of the Ritz projection z_h of z . This may be even more simplified to the cell-error indicators

$$\eta_K^{(3)} = c_I h_K^{3/2} \rho_K(u_h) \| [\partial_n z_h] \|_{\partial K}.$$

For the corresponding error estimator

$$E^{(3)}(u_h) := c_I \sum_{K \in \mathbb{T}_h} h_K^{3/2} \rho_K \| [\partial_n z_h] \|_{\partial K},$$

we usually observe strong over-estimation, i.e. $I_{\text{eff}}^{(3)} \gg 1$ (see Table 4.1) depending on what value we set for the interpolation constant c_I .

Approximation by local residual problems

On each cell K , we solve the local Neumann problem (see Bank and Weiser [17])

$$(\nabla v_K, \nabla \psi_h)_K = (R_h, \psi_h)_K + (r_h, \psi_h)_{\partial K} \quad \forall \, \psi_h \in V_K,$$

where $V_K = \{q \in \tilde{Q}_2(K),\, q \perp \tilde{Q}_1(K)\}$. Then, in view of the relation

$$\begin{aligned} |(\nabla v_K, \nabla (z - I_h z))_K| &\leq \|\nabla v_h\|_K \|\nabla (z - I_h z))_K\| \\ &\leq c_I h_K \|\nabla v_h\|_K \|\nabla^2 z\|_K \\ &\approx c_I h_K^{1/2} \|\nabla v_h\|_K \| [\partial_n z] \|_{\partial K}, \end{aligned}$$

the local error indicators may be defined as

$$\eta_K^{(4)} = c_I h_K^{1/2} \|\nabla v_K\|_K \|[\partial_n z_h]\|_{\partial K}.$$

In this way, one obtains fairly good bounds for the error in the energy norm. However, for local error functionals the results are not better than with the other simplified estimators. In such cases the crucial point seems more the accuracy in the approximation of the dual solution than in the evaluation of the residuals of u_h (see Backes [13]).

Numerical test

The effectivity of the first three of these error estimators has been tested for the 2-D model problem (4.3) with the solution $u(x) = (1-x_1^2)(1-x_2^2)\sin(4x_1)\sin(4x_2)$ for the two output functionals

$$J_1(u) := |S|^{-1} \int_S u\,dx, \quad S := [-\tfrac{1}{2}, 0] \times [0, \tfrac{1}{2}],$$
$$J_2(u) := u(\tfrac{1}{2}, \tfrac{1}{2}).$$

Table 4.1 shows the corresponding effectivity indices obtained on sequences of locally refined meshes on the basis of the local error indicators $\eta_K^{(i)}$, $i = 1, 2, 3$. It turns out that the cheap interpolation-based estimator $E^{(2)}(u_h)$ is almost as effective as the more expensive estimator $E^{(1)}(u_h)$. Therefore, in the following, we will almost exclusively use the first one in the presented numerical tests.

Table 4.1: *Effectivity of weighted error indicators for the mean error $J_1(e)$ (left) and the point-error $J_2(e)$ (right); from Richter [123].*

N	$J_1(e)$	$I_{\text{eff}}^{(1)}$	$I_{\text{eff}}^{(2)}$	$I_{\text{eff}}^{(3)}$
81	$7.6{\cdot}10^{-2}$	1.01	1.05	393
151	$2.7{\cdot}10^{-2}$	1.00	1.07	337
653	$3.6{\cdot}10^{-3}$	0.99	0.99	229
1435	$1.4{\cdot}10^{-3}$	1.00	1.00	177
2937	$6.5{\cdot}10^{-4}$	1.00	0.98	120
6249	$3.2{\cdot}10^{-4}$	1.00	1.00	84
12995	$1.2{\cdot}10^{-4}$	1.00	0.99	84
26603	$7.2{\cdot}10^{-5}$	1.00	1.00	55

N	$J_2(e)$	$I_{\text{eff}}^{(1)}$	$I_{\text{eff}}^{(2)}$	$I_{\text{eff}}^{(3)}$
81	$4.2{\cdot}10^{-1}$	0.17	0.18	69
151	$1.4{\cdot}10^{-1}$	0.26	0.20	83
635	$1.0{\cdot}10^{-2}$	0.30	0.28	123
1443	$2.9{\cdot}10^{-3}$	0.39	0.40	129
2875	$9.4{\cdot}10^{-4}$	0.52	0.51	130
6229	$2.8{\cdot}10^{-4}$	0.61	0.60	147
12521	$1.0{\cdot}10^{-4}$	0.74	0.72	148
26903	$3.9{\cdot}10^{-5}$	0.82	0.80	155

Remark 4.5. The traditional energy-norm error estimator η_E is known to be not only *reliable*, i.e., it provides a safe upper bound for the energy error, but also ***efficient*** , i.e., it is asymptotically sharp in the sense that

$$c_1 \|\nabla e\| \le \eta_E \le c_2 \big\{ \|\nabla e\| + \|f - \bar{f}_h\| \big\},$$

where $\bar{f}_h$ is the piecewise constant interpolation of f (see, e.g., Verfürth [132]). This means that the estimator is up to a constant asymptotically correct. A corresponding result is not possible in general for estimators η_ω of locally defined error quantities. Already the transition from the *error representation* to the *error estimate* in terms of (non-negative) cell-wise error indicators is critical, since by this localization the asymptotic sharpness of the global error representation may get lost. To illustrate this, consider the case $J(u) = u(0)$ and assume that the exact as well as the approximate solution are anti-symmetric with respect to the x_1-axis. Then, $e(0) = 0$, but usually $\sum_{K\in\mathbb{T}_h} \eta_K \neq 0$.

4.2 Mesh adaptation

Next, we address the practical aspects of successive mesh adaptation. Suppose that on the meshes $\mathbb{T}_h$, we have local error indicators η_K extracted from an a posteriori error estimate

$$|J(e)| \leq \eta := \sum_{K\in\mathbb{T}_h} \eta_K, \qquad N := \#\{K \in \mathbb{T}_h\}. \tag{4.4}$$

Using this information the computational mesh may be adapted using various different strategies. For quadrilateral meshes, as considered here, the refinement and coarsening is facilitated by using *hanging nodes*. The global conformity of the finite element ansatz is preserved since the unknowns at hanging nodes are eliminated by interpolation between the neighboring 'regular' nodes.

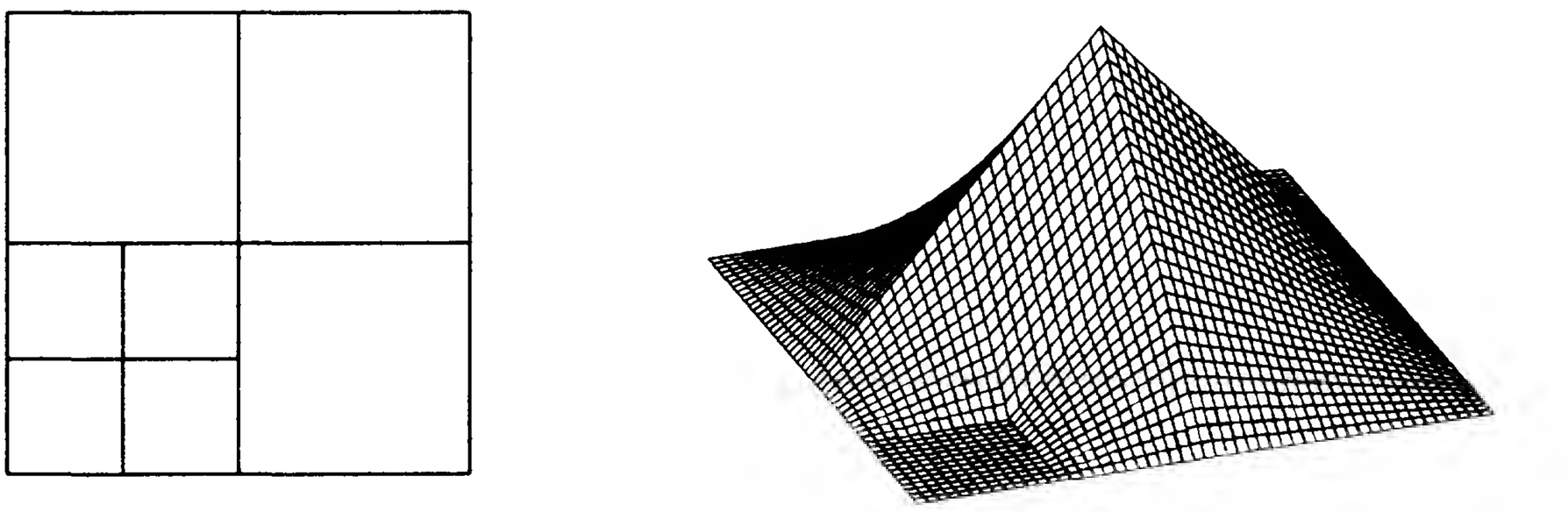


Figure 4.1: Q_1 *nodal basis function on a patch of cells with hanging nodes*

We note that there are several alternative strategies for realizing local mesh refinement. The occurrence of hanging nodes can be avoided by using special 'transition cells' (triangles or quadrilaterals) which bridge from cells of width h to those of width $h/2$. The construction of such cells may be complicated in 3-D. It may also cause a spreading of the refinement zone which implies extra work for de-refining. However, it is basically a question of taste which technique of

mesh organization one prefers. Refinement and coarsening of quadrilateral meshes involving the use of hanging nodes proceeds as indicated in Figure 4.2.

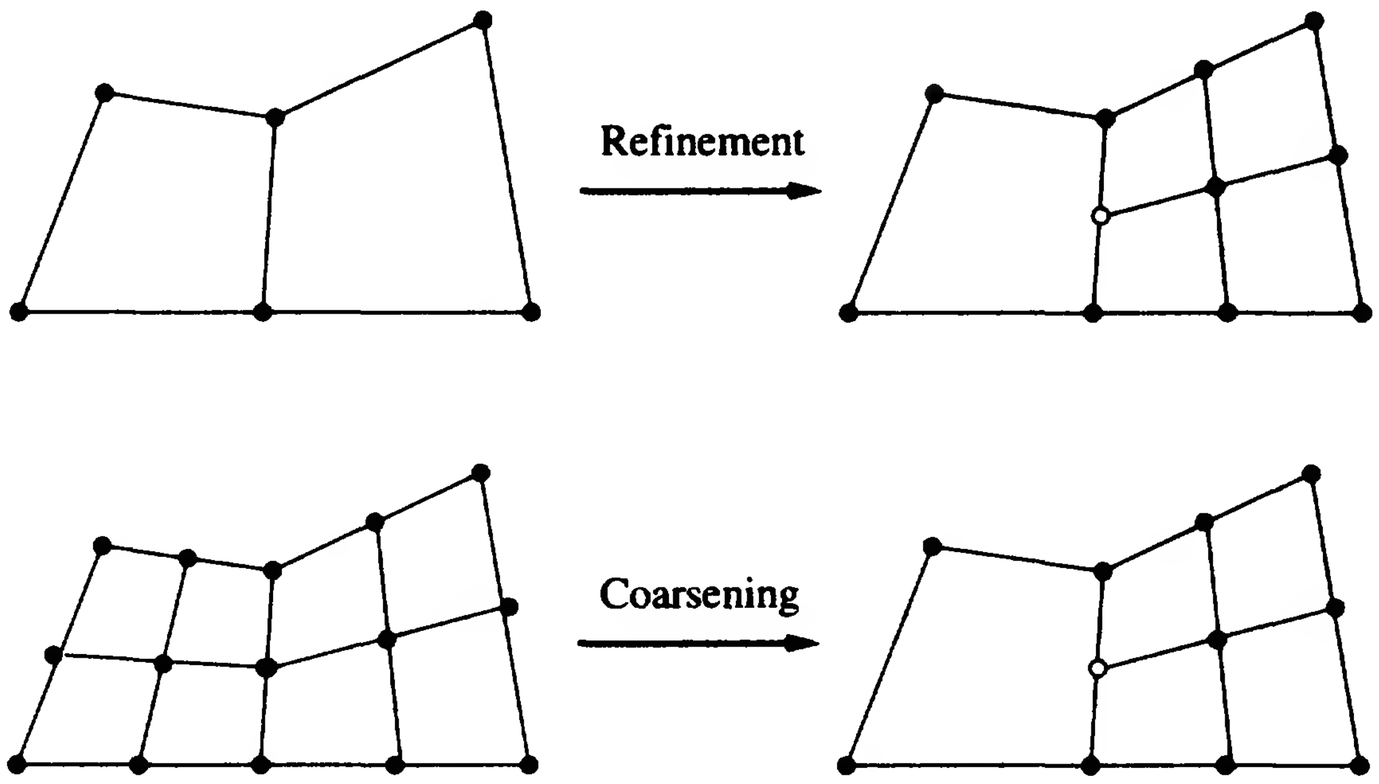


Figure 4.2: *Refinement and coarsening in quadrilateral meshes.*

At first, we have to check whether on the current mesh $\mathbb{T}_h$ the *stopping criterion*

$$\eta \leq TOL$$

is already satisfied. If this is the case, then u_h is accepted as approximation to u that represents the target quantity $J(u)$ by $J(u_h)$ within the desired tolerance TOL. Otherwise, the next refinement cycle is started. To this end, the cells in the current mesh are ordered according to

$$\eta_{K_1} \geq \cdots \geq \eta_{K_i} \geq \cdots \geq \eta_{K_N}.$$

Then, the mesh adaptation may be organized using one of the following strategies.

Error-balancing strategy

Below, we will give an argument which indicates that an *optimal* mesh is characterized by equilibrated error indicators, such as

$$\eta_{K_i} \approx \frac{TOL}{N}, \quad i = 1, \ldots N,$$

which implies $\eta(u_h) \approx TOL$. However, the so-called *error-balancing strategy* based on this idea is 'implicit' as it involves the current number of mesh cells N which is obtained only at the end of the adaptation cycle. We check, starting from $i = 1, j = 0$, whether

$$\eta_{K_i} \leq \frac{TOL}{N + 3j}.$$

If this is not satisfied, then the cell K_i is refined, the counters j and i are increased by one, and one proceeds to the next smaller η_{K_i}. But, if the condition is satisfied, then the new mesh $\mathbb{T}_h^{new}$ is reached. This strategy is potentially optimal but involves many expensive checking operations and is therefore impracticable.

Fixed-error-reduction or fixed-rate strategy

For fractions X, Y, with $1-X > Y$, determine indices $N_*, N^* \in \{1,\dots,N\}$, such that

$$\sum_{i=1}^{N^*} \eta_{K_i} \approx X\eta, \qquad \sum_{i=N_*}^{N} \eta_{K_i} \approx Y\eta.$$

Then, the cells $K_1,\dots,K_{N^*}$ are refined and the cells $K_{N^*},\dots,K_N$ are coarsened. Common choices are $X = 0.2$ and $Y = 0.1$. Alternatively, in the *fixed-rate strategy*, one refines $X\cdot N$ and coarsens $Y\cdot N$ cells with largest and smallest error indicators, respectively. For appropriate choices of X, Y this accomplishes to keep the number of cells almost constant in the course of the mesh adaptation process.

Mesh-optimization strategy

The information contained in the error representation (4.1) may be used directly to construct an 'optimal' mesh on which the error tolerance $\eta \approx \mathrm{TOL}$ is achieved, i.e. skipping the intermediate 'one-level' refinement steps. Here, a mesh $\mathbb{T}_h = \{K\}$ is characterized by a continuous *mesh-size function* $h(x)$, where $h_{|K} \approx h_K$. Further, it is assumed that the error estimator is related to a continuous limit of the form

$$\sum_{K\in\mathbb{T}_h} \eta_K \approx \int_\Omega h(x)^2 \Phi(x)\,dx =: \eta(h),$$

with $\Phi := (\Phi_1+\Phi_2)\Phi_3$ a mesh-independent weighting function. We note that the latter requirement rules out cases in which the functional $J(\cdot)$ and hence also the dual solution implicitly depend on the mesh size, as for example in the estimation of the error in the energy or the L^2 norm (cf. Section 3.2). The components of Φ are defined by the limiting processes of residuals and weights, for $\mathrm{TOL} \to 0$:

$$\max_{x\in K} |f+\Delta u_h| \approx \max_{x\in K} \Phi_1, \qquad \tfrac{1}{2} h_K^{-1} \max_{x\in K} |[\partial_n u_h]_{\partial K}| \approx \max_{x\in K} \Phi_2, \tag{4.5}$$

$$h_K^{-2} \max_{x\in K} |z - I_h z| \approx \max_{x\in K} \Phi_3. \tag{4.6}$$

Here, the mesh-size power $h(x)^2$ is related to the 'order' of the considered finite element ansatz. The justification of these assumptions will be discussed in the next chapter. The function $\Phi(x)$ may have strong (regularized) singularities which will possibly require the mesh-size $h(x)$ to reduce down towards zero even for tolerance $TOL > 0$. Further, we introduce the mesh complexity formula

$$N = \sum_{K\in\mathbb{T}_h} h_K^d h_K^{-d} \approx \int_\Omega h(x)^{-d}\,dx =: N(h),$$

which denotes the number of degrees of freedom for a mesh-size function $h(x)$. With this notation, we obtain the following result.

Proposition 4.6. *The mesh-optimization problem*

$$\eta(h) \to \min, \quad N(h) \le N_{\max} \tag{4.7}$$

is solved by

$$h_{\mathrm{opt}}(x) = \Big(\frac{W}{N_{\max}}\Big)^{1/d} \Phi(x)^{-1/(2+d)}, \tag{4.8}$$

provided that

$$W := \int_\Omega \Phi(x)^{d/(2+d)}\, dx < \infty\,.$$

For an 'optimal' mesh, there hold

$$TOL = \frac{W^{(2+d)/d}}{N^{2/d}} \quad \textit{and} \quad N \approx \frac{W^{(2+d)/2}}{TOL^{d/2}}. \tag{4.9}$$

Proof. Following the classical Lagrange approach, we introduce the Lagrangian

$$L(h,\lambda) = \eta(h) + \lambda\{N(h) - N_{\max}\}$$

with Lagrangian multiplier $\lambda \in \mathbb{R}$. Then, the optimal mesh-size function h_{opt} is characterized as stationary point of the first-order optimality condition

$$\frac{d}{dt} L(h+t\varphi,\lambda)_{|t=0} = 0, \quad \frac{d}{dt} L(h,\lambda+t\mu)_{|t=0} = 0,$$

for all admissible variations φ and μ. This means that

$$2h(x)\Phi(x) - d\lambda h(x)^{-d-1} = 0, \qquad \int_\Omega h(x)^{-d}\, dx - N_{\max} = 0,$$

and, consequently,

$$h(x) = \Big(\frac{2}{d\lambda}\Phi(x)\Big)^{-1/(2+d)}, \qquad \Big(\frac{2}{d\lambda}\Big)^{d/(2+d)} \int_\Omega \Phi(x)^{d/(2+d)}\, dx = N_{\max}.$$

From this, we deduce the desired relations

$$\lambda \equiv \frac{2}{d}\, h(x)^{2+d}\Phi(x)\,, \qquad h_{\mathrm{opt}}(x) = \Big(\frac{W}{N_{\max}}\Big)^{1/d} \Phi(x)^{-1/(2+d)}.$$

From the formula for h_{opt}, we conclude that on the optimal mesh, there holds

$$TOL = \Big(\frac{W}{N}\Big)^{2/d} \int_\Omega \Phi(x)^{-2/(2+d)}\Phi(x)\, dx = \frac{W^{(2+d)/d}}{N^{2/d}},$$

which proves (4.9). □

Remark 4.7. We note that in two dimensions the optimal mesh complexity is

$$TOL = \mathcal{O}(N^{-1}) \quad \text{or} \quad N = \mathcal{O}(TOL^{-1}),$$

for linear or bilinear finite elements, provided that $\sup_{TOL\to 0} W < \infty$ is satisfied. This is the case even for rather 'irregular' functionals $J(\cdot)$. For example, the evaluation of $J(u) = \partial_i u(a)$ leads to $\Phi(x) \approx (|x-a|^2 + TOL^2)^{-3}$ and, consequently,

$$\sup_{TOL\to 0} W \approx \int_\Omega |x-a|^{-3/2}\, dx < \infty.$$

In the case that $\sup_{TOL} W = \infty$, as for example for $J(u) = \partial_i^2 u(a)$, the optimal mesh complexity becomes

$$TOL = \mathcal{O}(N^{\alpha-1}) \quad \text{or} \quad N = \mathcal{O}(TOL^{-\alpha-1}),$$

with some $\alpha > 0$. In particular, for the latter functional, we easily find $TOL = \mathcal{O}(N^{-1}|\log(N)|)$.

Remark 4.8. The result of Proposition 4.6 implies that the optimal mesh-size distribution is characterized by the *equilibration property*

$$h(x)^{2+d}\Phi(x) \equiv \left(\frac{TOL}{W}\right)^{(2+d)/d} = \text{const.} \tag{4.10}$$

This justifies the strategy of equilibrating the local error indicators η_K,

$$\eta(h) = \int_\Omega h(x)^2\Phi(x)\, dx \approx \sum_{K\in\mathbb{T}_h} h_K^{2+d}\Phi_K = \sum_{K\in\mathbb{T}_h} \eta_K,$$

as used in the *error-balancing strategy.*

Let the weight function $\Phi(x)$ be bounded. Then, once a balanced mesh satisfying (4.10) is reached, a maximum increase of accuracy is achieved by uniform mesh refinement. If Φ is singular, then the 'optimal' mesh size tends to zero at these singularities even for $TOL > 0$.

Remark 4.9. Alternatively to (4.7), we can also consider the mesh-optimization problem

$$N(h) \to \min, \quad \eta(h) \le TOL,$$

which has the solution

$$h_{\text{opt}}(x) = \left(\frac{TOL}{W}\right)^{1/2} \Phi(x)^{-1/(2+d)}.$$

Remark 4.10. Although the mesh-optimization strategy seems very attractive, its realization involves several problems:

- The derivation of the formula for an optimal mesh-size distribution is based on the assumption that on the considered meshes the cell-residuals behave in an optimal way under refinement, i.e., $\rho_K \approx h_K$. This is hard to prove and may not be true in general; see the discussion of this question in Chapter 5.
- The numerical approximation of the weighting function $\Phi(x)$ should provide more information than can be cheaply obtained using only information on the current mesh.
- The explicit formulas for $h_{\text{opt}}(x)$ have to be used with care in designing a mesh as their derivation implicitly assumes that they actually correspond to *scalar* mesh-size functions of *isotropic* meshes, a condition, however, which is not incorporated into the formulation of the mesh-optimization problems.

A detailed study of how to utilize mesh optimization for the model problem has been made by Richter [123].

Remark 4.11. *(The role of regularization)* At the beginning of this chapter, we had mentioned that in the case of a 'singular' functional like, for example, that for evaluation of the derivative point value, regularization is necessary,

$$J_\varepsilon(u) := |B_\varepsilon|^{-1} \int_{B_\varepsilon} \partial_1 u \, dx,$$

To demonstrate the effect caused by regularizing with $\varepsilon = TOL$, as proposed above, compared to $\varepsilon = h_{min}$, we show the results of a numerical test for the model situation considered in Example 3.4, see Table 4.2. We see that in this case regularization drastically reduces the number L of refinement steps for reaching the same accuracy level which means less computational work.

Table 4.2: *Results for computing $\partial_1 u(0)$ in Example 3.4 using regularization with $\varepsilon = TOL$ (right) and $\varepsilon = h_{min}$ (left); from Becker and Rannacher [30].*

	$\varepsilon = h_{min}$			$\varepsilon = TOL$		
TOL	N	L	$\lvert\partial_1 e(0)\rvert$	N	L	$\lvert\partial_1 e(0)\rvert$
1	40	4	$2.57{\cdot}10^{-0}$	40	4	$2.57{\cdot}10^{-0}$
4^{-1}	124	6	$7.38{\cdot}10^{-1}$	64	4	$1.47{\cdot}10^{-0}$
4^{-2}	448	11	$1.35{\cdot}10^{-3}$	148	6	$7.51{\cdot}10^{-1}$
4^{-3}	1780	15	$1.19{\cdot}10^{-3}$	940	9	$4.10{\cdot}10^{-1}$
4^{-4}	6328	19	$2.20{\cdot}10^{-3}$	4912	12	$4.14{\cdot}10^{-3}$
4^{-5}	25984	24	$4.28{\cdot}10^{-4}$	20980	15	$2.27{\cdot}10^{-4}$
4^{-6}	95260	28	$1.39{\cdot}10^{-4}$	86740	17	$5.82{\cdot}10^{-5}$

4.3 Use of error estimators for post-processing

The duality approach described so far for estimating the discretization error can also be used for increasing the accuracy in the approximation of the target quantity. Consider the model situation as before, i.e. the Poisson problem in 2-D written in variational form as

$$a(u,\psi) = (f,\psi) \quad \psi \in V, \tag{4.11}$$

and discretized by

$$a(u_h,\psi_h) = (f,\psi_h) \quad \psi_h \in V_h. \tag{4.12}$$

The output functional is $J(\cdot)$. Let $z \in V$ be the corresponding dual solution and $z_h \in V_h$ its Ritz projection on the 'primal' mesh $\mathbb{T}_h$. By definition, there holds

$$J(u) = a(u,z) = (f,z), \tag{4.13}$$

i.e., if z were known, the target quantity $J(u)$ is determined solely by the data f. This observation will be used below for post-processing the approximation $J(u_h)$. For this, we recall the identity

$$J(u) = J(u_h) + a(e,z) = J(u_h) + \rho(u_h)(z-z_h),$$

where

$$\rho(u_h)(z-z_h) = (f,z-z_h) - a(u_h,z-z_h).$$

Above, we have discussed ways of constructing approximations to the dual solution z, e.g. the patch-wise biquadratic interpolation $\tilde{z}_h := I_{2h}^{(2)} z_h$ of z_h on the mesh $\mathbb{T}_h$. This led us to the approximate error representation

$$J(e) \approx \rho(u_h)(\tilde{z}_h - z_h) = \rho(u_h)(\tilde{z}_h).$$

Rewriting this relation as

$$J(u) \approx \tilde{J}_1(u_h) := J(u_h) + (f,\tilde{z}_h) - a(u_h,\tilde{z}_h),$$

we obtain a presumably better approximation to $J(u)$ than is $J(u_h)$. The error can be written as

$$J(u) - \tilde{J}_1(u_h) = J(e) - \rho(u_h)(\tilde{z}_h) = a(e,z) - a(u_h,\tilde{z}_h) = a(e,z-\tilde{z}_h),$$

which implies

$$|J(u) - \tilde{J}_1(u_h)| \le \|\nabla e\| \, \|\nabla(z-\tilde{z}_h)\|. \tag{4.14}$$

Since it is not clear whether $\tilde{z}_h$ is a reasonably better approximation to z than z_h, this estimate is of only questionable value. Further, the two energy-norm errors correspond both to the 'primal' mesh $\mathbb{T}_h$ and can therefore not be minimized independently. It would be desirable to have the possibility of using independent meshes $\mathbb{T}_h$ for u_h and $\mathbb{T}_h^*$ for z_h in constructing an approximation of $J(u)$. This is accomplished in the following proposition (see Richter [123]).

Proposition 4.12. *Let $\mathbb{T}_h$ and $\mathbb{T}_h^*$ be two independent meshes and V_h and V_h^* corresponding finite element spaces in which the Ritz projections u_h and z_h^* of u and z are computed. Further, denote by $\tilde{z}_h^*$ the patch-wise biquadratic interpolation of z_h^* on the dual mesh $\mathbb{T}_h^*$. Then, for the post-processed approximation*

$$\tilde{J}_2(u_h) := J(u_h) + (f, \tilde{z}_h^*) - a(u_h, \tilde{z}_h^*),$$

there holds the estimate

$$|J(u) - \tilde{J}_2(u_h)| \leq \|\nabla e\| \, \|\nabla(z - \tilde{z}_h^*\|. \tag{4.15}$$

Proof. We have

$$\begin{aligned} J(u) - \tilde{J}_2(u_h) &= J(u) - J(u_h) - \rho(u_h)(\tilde{z}_h^*) \\ &= (f, z) - a(u_h, z) - (f, \tilde{z}_h^*) + a(u_h, \tilde{z}_h^*) \\ &= (f, z - \tilde{z}_h^*) - a(u_h, z - \tilde{z}_h^*) \\ &= a(e, z - \tilde{z}_h^*). \end{aligned}$$

This implies the assertion. □

We emphasize that in the estimate (4.15) the two energy-norm terms can be minimized independently by optimizing the primal and dual meshes $\mathbb{T}_h$ and $\mathbb{T}_h^*$, respectively. One may hope to obtain an even better approximation by

$$\tilde{J}_3(u_h) := \tilde{J}_2(\tilde{u}_h) = J(\tilde{u}_h) + (f, \tilde{z}_h^*) - a(\tilde{u}_h, \tilde{z}_h^*),$$

where $\tilde{u}_h$ is the patch-wise biquadratic interpolation of u_h. Below, we will test all three post-processed approximations to $J(u)$ for the model problem on $\Omega = (-1,1)^2$ with the prescribed solution $u(x) = (1-x_1^2)(1-x_2^2)\exp(1-x_2^{-4})$ and the output functional

$$J(u) = \int_{-1}^{1} u(x_1, 0.5)\, dx.$$

In this example primal and dual solution have irregularities at different locations such that it is to expected that maximal efficiency is achieved using different meshes for u_h and z_h as shown in Figure 4.3.

Figure 4.4 shows the mesh efficiencies of $J(u_h)$ and the three post-processed approximations $\tilde{J}_1(u_h)$, $\tilde{J}_2(u_h)$, and $\tilde{J}_3(u_h)$. The mesh refinements are driven by energy-norm error indicators as derived in Section 3.2 separately on the primal and dual meshes $\mathbb{T}_h$ and $\mathbb{T}_h^*$, respectively. We see that $\tilde{J}_1(u_h)$ indeed does not bring significant advantages over the original approximation $J(u_h)$. The two other approximations $\tilde{J}_2(u_h)$ and $\tilde{J}_3(u_h)$ which use different meshes for u and z are clearly superior and show a mesh complexity like $TOL \approx N^{-2}$ ($N = N_{\text{primal}} + N_{\text{dual}}$). This is to be compared to the optimal complexity $TOL \approx N^{-1}$ obtained above in Proposition 4.12.

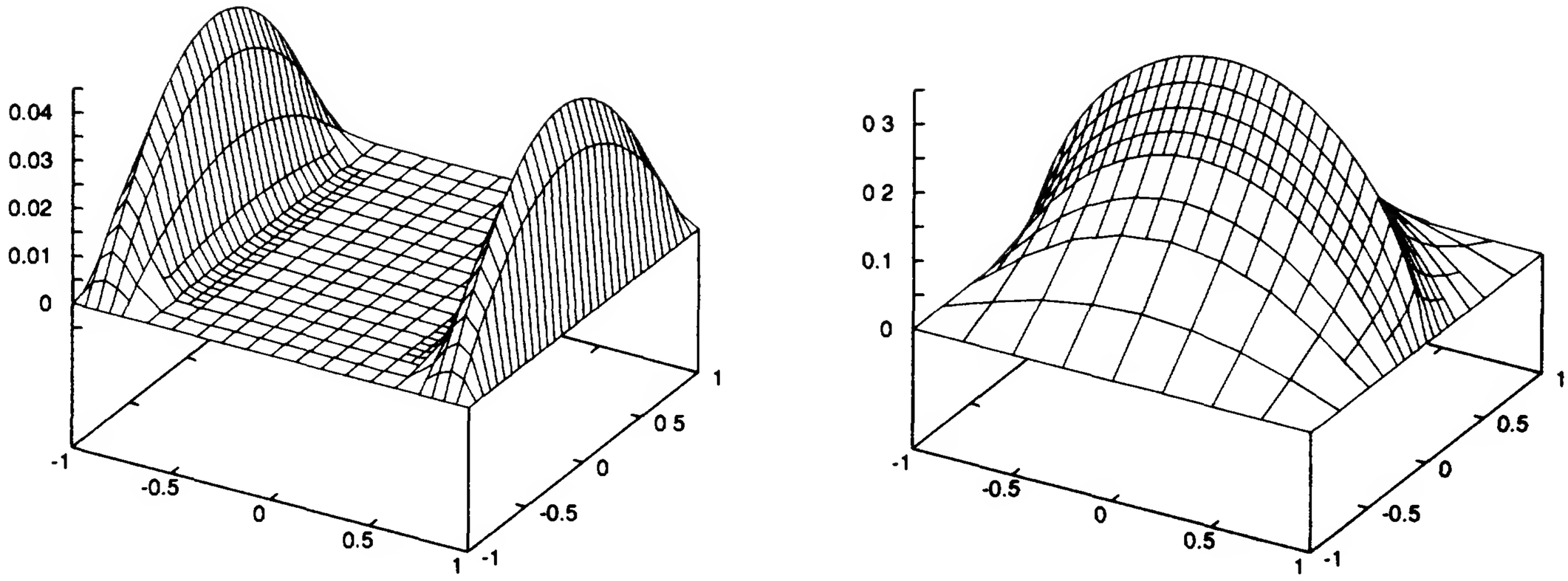


Figure 4.3: *Primal (left) and dual (right) solution of the model problem; from Richter [123].*

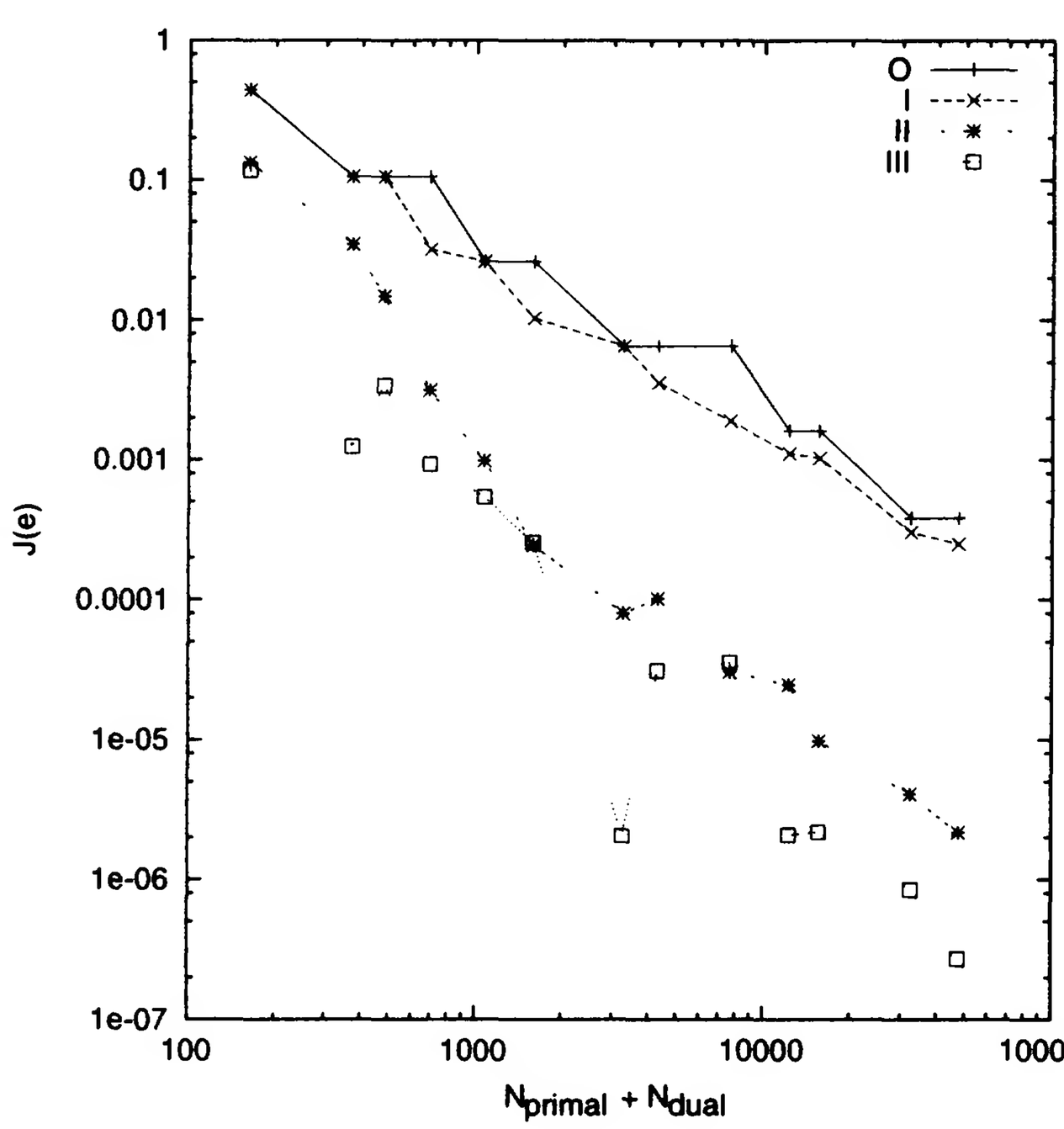


Figure 4.4: *Mesh efficiencies of post-processing:* $J(u_h)$ *(solid line, symbol +),* $\tilde{J}_1(u_h)$ *(broken line, symbol ×),* $\tilde{J}_2(u_h)$ *(broken line, symbol ∗),* $\tilde{J}_3(u_h)$ *(dotted line, symbol □); from Richter [123].*

4.4 Towards anisotropic mesh adaptation

Sometimes *isotropic* mesh refinement as discussed so far is not efficient for properly resolving direction-dependent features of the solution. For example, in singularly perturbed problems of the form

$$-\varepsilon\Delta u + bu = f \ \text{ in } \Omega, \qquad u_{|\partial\Omega} = 0,$$

with small coefficient ε, boundary layers may occur in which the solution has a large derivative in normal direction to the boundary while it varies only slowly in tangential direction. A similar phenomenon occurs in 3D along reentrant edges of the domain (edge singularities). In such a situation it is appropriate to use meshes which are anisotropically refined in the sense that the cells along the boundary are much thinner in normal than in tangential direction, see Figure 4.5.

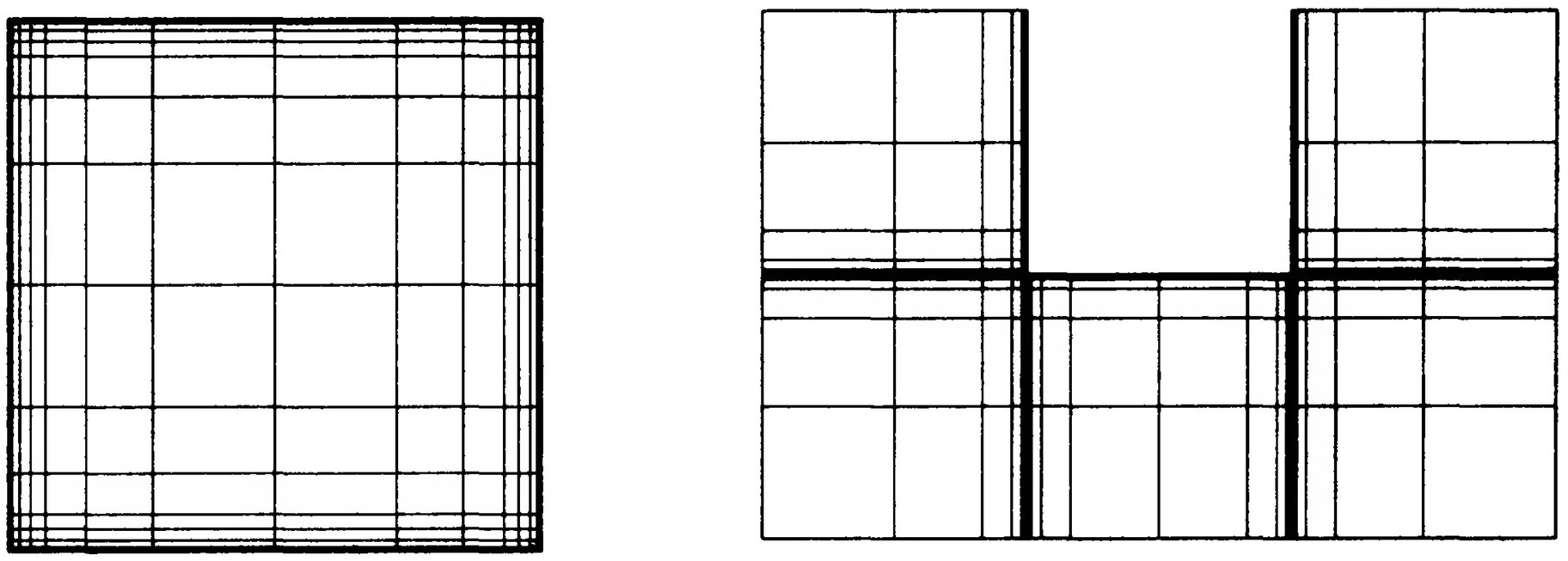


Figure 4.5: *Locally anisotropic tensor-product meshes.*

The questions are now whether the weighted error estimator contains information about anisotropy in the exact solution, and whether we can extract local indicators which tell us how to adapt the mesh according to (i) orientation of cells, and (ii) optimal stretching of cells. This aspect of automatic mesh adaptation is a rather difficult one and still the subject of current research.

To fix ideas, consider a given function $u \in C^2(\bar{B}_1)$ on the ball $B_1 = \{x \in \mathbb{R}^d, |x| < 1\}$ and determine the direction of smallest line-wise L^2 error of linear interpolation, i.e., find the unit vector e such that

$$\int_\Gamma |u - I_h u|^2 \, ds \approx \int_\Gamma |\partial_e^2 u|^2 \, ds = \int_\Gamma |(H(u)e, e)| \, ds \ \to \ \min_\Gamma .$$

Here, $\Gamma = \{x \in B_1, x = se, 0 \le s < 1\}$, and $H(u) := \nabla^2 u$ is the (symmetric) Hessian matrix of u. Hence, the interpolation error becomes minimal for $e = e_{\min}$ being the eigenvector corresponding to the eigenvalue of $H(u)$ with minimal modulus. We see that all the information about the best orientation of cells for minimizing the interpolation error is available by the Hessian matrix. The eigenvalue

quotient $\sigma = |\lambda_{\max}/\lambda_{\min}|$ determines the cell aspect ratio and the corresponding (orthogonal) eigenvectors $e_{\min}$ and $e_{\max}$ the orientation of the cell.

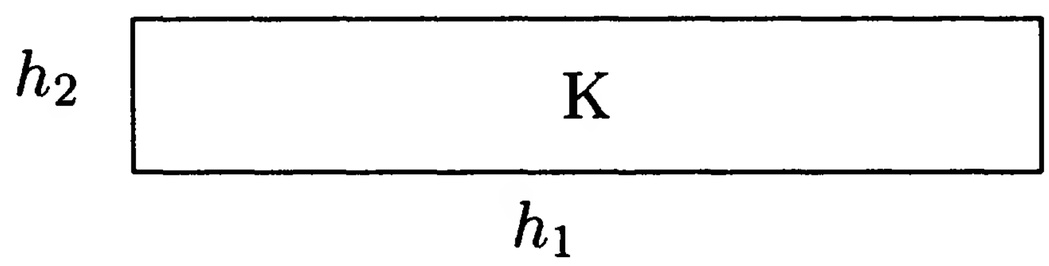


Figure 4.6: *Cell of a Cartesian mesh.*

Quadrilateral meshes are not very suited for dynamical cell reorientation. For this purpose, triangular meshes are more flexible. Therefore, we will consider adaptive cell stretching alone and, for simplicity, will concentrate on the construction of 'optimal' Cartesian meshes consisting of cells K with area $|K| = h_1 h_2$, as shown in Figure 4.6. Again 'hanging' nodes are allowed. The maximum $\sigma_K := \max\{h_1/h_2, h_2/h_1\}$ is the *cell aspect ratio* and $\sigma_h := \max_{K\in\mathbb{T}_h} \sigma_K$ the *maximum mesh aspect ratio.*

We recall the a posteriori error representation

$$J(e) = \sum_{K\in\mathbb{T}_h} \left\{(f+\Delta u_h, z-I_h z)_K + \tfrac{1}{2}([\partial_n u_h], z-I_h z)_{\partial K\setminus\partial\Omega}\right\}. \tag{4.16}$$

Practical experience and theoretical analysis show that on *isotropic* meshes the edge-residual terms $([\partial_n u_h], z-I_h z)_{\partial K}$ can be made to dominate the cell-residual terms $(f+\Delta u_h, z-I_h z)_K$, see Exercise 5.2. Let us assume that the same is true also for anisotropic meshes (for some theoretical support see Kunert and Verfürth [99]). Therefore, we only consider the edge-residual terms and as before suppose the approximation

$$h_i^{-1}[\partial_i u_h] \approx \partial_i^2 u \quad (i=1,\ldots,d).$$

Then, assuming again the second-order derivatives of u as constant, we obtain

$$\begin{aligned} |([\partial_n u_h], z-I_h z)_{\partial K}| &\approx h_1^3 h_2 |\partial_2^2 u|\,|\partial_1^2 z| + h_1 h_2^3 |\partial_1^2 u|\,|\partial_2^2 z| \\ &= |K|\left\{h_1^2|\partial_2^2 u|\,|\partial_1^2 z| + |K|^2 h_1^{-2}|\partial_1^2 u|\,|\partial_2^2 z|\right\}. \end{aligned}$$

Minimizing this with respect to h_1 yields the necessary condition

$$2h_1|\partial_2^2 u|\,|\partial_1^2 z| - 2|K|^2 h_1^{-3}|\partial_1^2 u|\,|\partial_2^2 z| = 0 \quad\Rightarrow\quad h_1^4 = |K|^2 \frac{|\partial_1^2 u|\,|\partial_2^2 z|}{|\partial_2^2 u|\,|\partial_1^2 z|},$$

and, consequently,

$$\frac{h_1}{h_2} \approx \left(\frac{|\partial_1^2 u|\,|\partial_2^2 z|}{|\partial_2^2 u|\,|\partial_1^2 z|}\right)^{1/2}. \tag{4.17}$$

This result is counterintuitive as it does not indicate the optimal cell stretching. To see this, consider the case that u is linear in x_1-direction, i.e., $\partial_1^2 u \equiv 0$, and that z is isotropic. Then, formula (4.17) would suggest to refine the cell in x_1 direction which is evidently the wrong decision. It seems that considering only the edge terms in (4.16) leads to contradictory results, and we rather have to take into account the whole combination of cell and edge residuals.

Remark 4.13. The development of a rigorous criterion for anisotropic mesh refinement on the basis of 'goal-oriented' error representations such as (4.16) must be left as an open problem. Here the difficulty is caused by the local interplay of the primal and dual solutions which may have significanty different regularity properties. In the special case of error estimation with respect to the energy-norm the dual solution coincides with the primal error such that both quantities behave very much the same way. Then, anisotropic mesh refinement can be surely guided by information from the jump residuals of u_h alone (see, e.g., Siebert [126]).

In view of the above discussion, we now follow a more heuristic approach and base the anisotropic cell adaptation on an estimate for the interpolation error. We recall the anisotropic interpolation error (see, e.g., Becker [18])

$$\|\nabla(u-I_h u)\|_K \leq c\left(h_1^2\|\partial_1\nabla u\|_K^2 + h_2^2\|\partial_2\nabla u\|_K^2\right)^{1/2}. \tag{4.18}$$

Hence, assuming the second-order derivatives as constant on K, we have

$$\|\nabla(u-I_h u)\|_K \leq c|K|^{1/2}\left(h_1^2|\partial_1\nabla u|^2 + |K|^2 h_1^{-2}|\partial_2\nabla u|^2\right)^{1/2}.$$

Minimizing this with respect to h_1 results in the necessary condition

$$2h_1|\partial_1\nabla u|^2 - 2|K|^2 h_1^{-3}|\partial_2\nabla u|^2 = 0 \quad\Rightarrow\quad h_1^2 = |K|\frac{|\partial_2\nabla u|}{|\partial_1\nabla u|},$$

and, consequently,

$$\frac{h_1}{h_2} \approx \frac{|\partial_2\nabla u|}{|\partial_1\nabla u|}.$$

In view of this result, we now consider the heuristic error indicator

$$\eta_K := \|\nabla(u-I_h u)\|_K\|\nabla(z-I_h z)\|_K,$$

which is minimized for

$$\frac{h_1}{h_2} \approx \frac{|\partial_2\nabla u|\,|\partial_2\nabla z|}{|\partial_1\nabla u|\,|\partial_1\nabla z|}. \tag{4.19}$$

This relation simultaneously reflects possible anisotropies in the primal and dual solution. The performance of anisotropic mesh adaptation based on the relations (4.17) and (4.19) will be compared in some numerical tests below.

Remark 4.14. We note that for the goal-oriented adaptation of purely tensor-product meshes an analogue of the mesh optimization strategy described in Section 4.2 can be developed, see Richter [123]. This approach yields meshes of optimal complexity in the case that the anisotropies of the primal and the dual solution are aligned with the coordinate directions. An extreme case occurs in Example 3.5, on a square domain, where the functional

$$J(u) = \int_{\partial\Omega} \partial_n u \, ds$$

results in a dual solution which is concentrated along the boundary $\partial\Omega$. Then, an optimal anisotropic refinement yields a mesh complexity of the order

$$TOL = \mathcal{O}(\sigma_h^{-1}),$$

for constant number N of cells. This is much better than the complexity $TOL = \mathcal{O}(N^{-1})$, which can be achieved with isotropic meshes.

Numerical test

As test case, we consider the Poisson problem on the square domain $\Omega = (-1,1)^2$,

$$-\Delta u = f \quad \text{in } \Omega, \quad u_{|\partial\Omega} = 0, \tag{4.20}$$

for the exact solution

$$u(x) = (1-x_1^2)^2(1-x_2)^2(kx_1^2+0.1)^{-1},$$

where the parameter $k = 1, 4, 16, 64, \ldots$, determines the strength of the anisotropy. The right hand side is determined as $f := -\Delta u$. This solution is shown for $k = 4$ and $k = 64$ in Figure 4.7. The quantity to be computed is the mean value

$$J(u) := |\Omega|^{-1} \int_\Omega u \, dx.$$

In this case the anisotropy is only in the primal solution while the dual solution satisfies $-\Delta z = 1$ and is isotropic.

The computation starts from a coarse uniform tensor-product mesh which is then successively adapted on the basis of the relations (4.17) ('Strategy I') and (4.19) ('Strategy II'). The efficiency of the resulting meshes is depicted in Figure 4.8. Apparently, the meshes produced by Strategy II on the basis of the heuristic relation (4.19) are more efficient as the anisotropy increases than those obtained by Strategy I on the basis of the 'rigorous' relation (4.17).

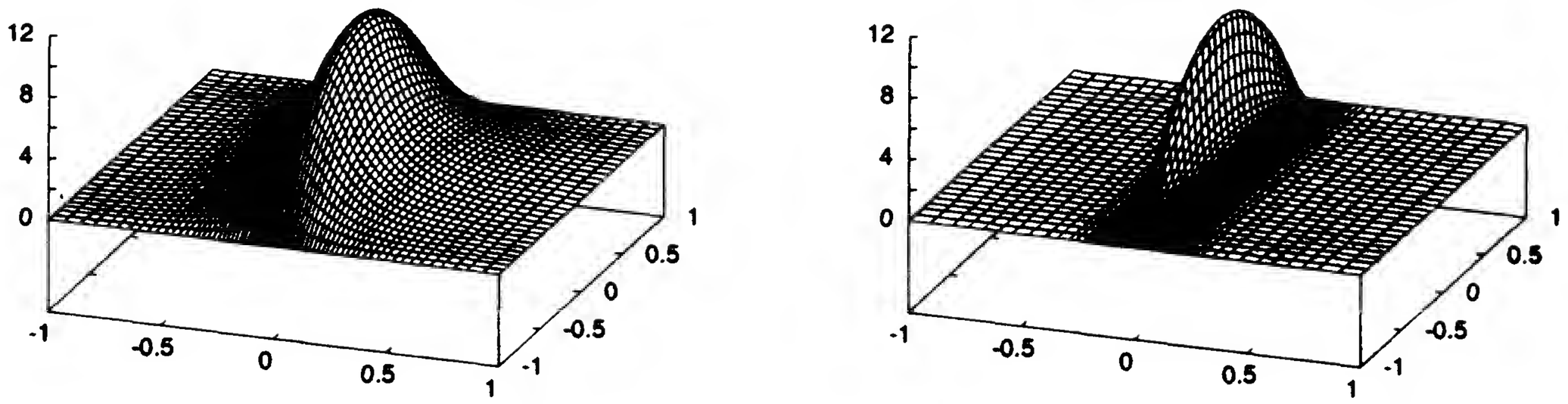


Figure 4.7: *Anisotropic solutions for* $k = 4$ *(left) and* $k = 64$ *(right); from Richter [123].*

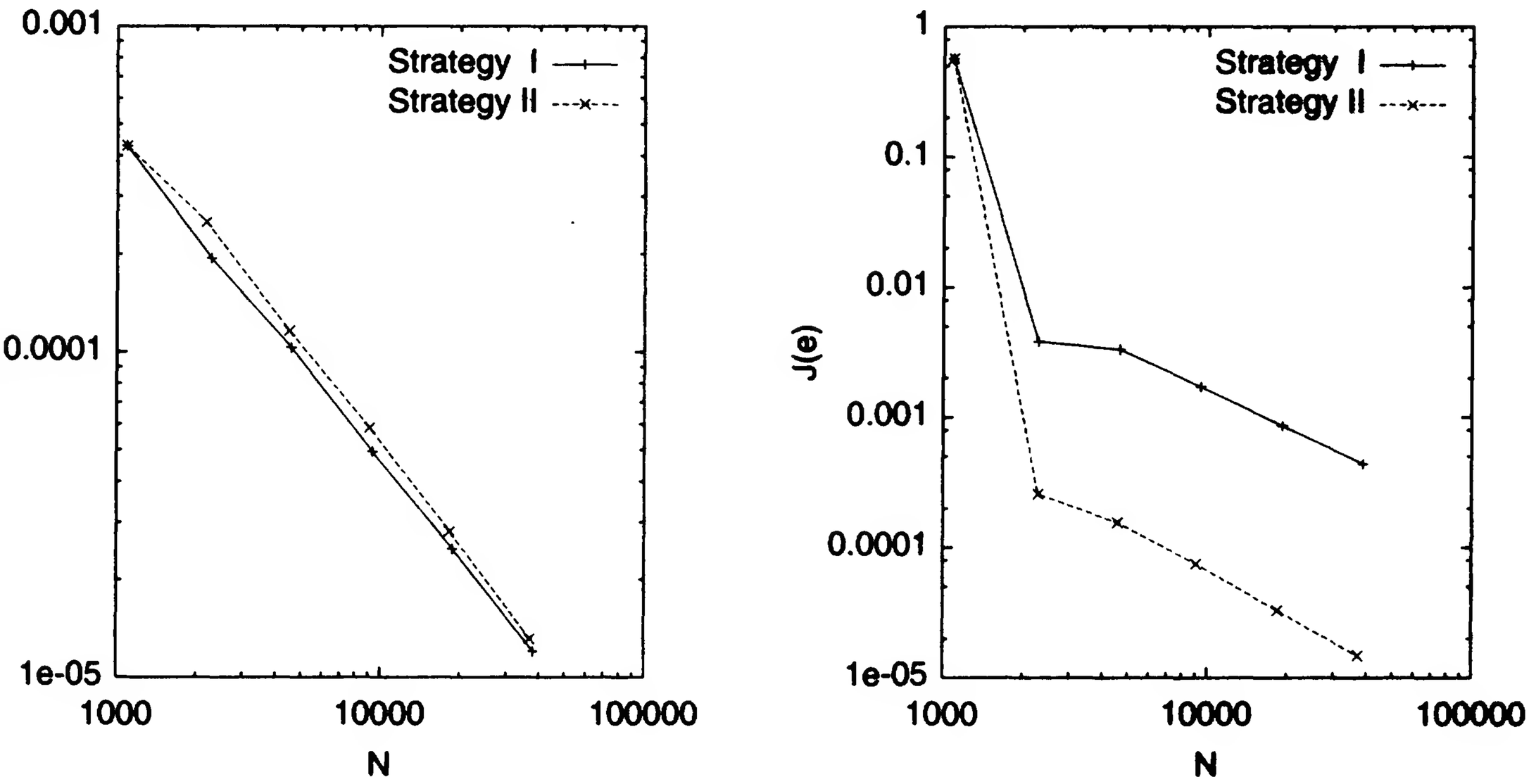


Figure 4.8: *Mesh efficiencies of anisotropic refinement by Strategy I and Strategy II, for* $k = 1$ *(left) and* $k = 64$ *(right); from Richter [123].*

4.5 Exercises

Exercise 4.1. The *fixed rate strategy* in a posteriori mesh adaptation refines and coarsens certain fractions X and Y of those cells with largest and smallest indicator values η_K, respectively,

$$\eta_{K_1} \geq \cdots \geq \eta_{K_i} \geq \cdots \geq \eta_{K_N}.$$

Design refining/coarsening strategies by specifying X, Y, such that

a) the number N of cells approximately doubles in each refinement cycle;

b) the number N of cells is approximately kept constant during the refinement process.

Exercise 4.2. The well-known Bramble-Hilbert theory guarantees the cell-wise error estimate

$$\|\nabla(u - I_h u)\|_K \leq c_I h_K \|\nabla^2 u\|_K$$

for the piecewise bilinear finite element nodal interpolation $I_h u$ on regular meshes, with a constant c_I independent of K. Sketch in 2D the argument that this estimate has an analogue on cells with one or two *hanging nodes*,

$$\|\nabla(u - I_h u)\|_K \leq c_I h_K \|\nabla^2 u\|_{\tilde{K}},$$

where $\tilde{K}$ is the mother cell of K on the preceding refinement level. Recall that at a hanging node the function value is set to be the average of the values at the two neighboring 'regular' nodes.

Exercise 4.3 *(Practical exercise).* Hanging nodes are commonly used to ease mesh refinement and coarsening in triangular or quadrilateral meshes. However, the presence of hanging nodes (and also transition cells) usually destroys the uniformity pattern of the mesh which may drastically reduce the accuracy of approximation. To demonstrate this, consider the model Poisson problem

$$-\Delta u = f \ \text{ in } \Omega, \quad u_{|\partial\Omega} = 0,$$

on $\Omega = (-1,1)^2$, with right-hand side $f(x) = \frac{1}{2}\pi^2 \cos(\frac{1}{2}x_1\pi)\cos(\frac{1}{2}x_2\pi)$ and corresponding exact solution $u(x) = \cos(\frac{1}{2}x_1\pi)\cos(\frac{1}{2}x_2\pi)$. Compute the point-value $\partial_1 u(0.5, 0.5)$ on a sequence of

a) uniform meshes with mesh-sizes $h = 2^{-1}$, $i = 1, 2, \ldots$;

b) locally refined meshes using the a posteriori error estimator and the corresponding refinement strategy of Exercise 2.3.

Compare the achieved accuracy in terms of the number of cells N.

Chapter 5

The Limits of Theoretical Analysis

In this chapter, we want to discuss some questions concerning the theoretical justification of the DWR method for goal-oriented mesh adaptivity presented so far. We will see that this task is rather demanding and poses several new questions for the theoretical analysis of the finite element method. In fact, relying on the available results from the literature, we do not reach very far, yet. Since several not very practical assumptions will be used, we dispense with stating formal propositions.

To illustrate the problem, let us consider the special case of the evaluation of the derivative of a smooth solution $u \in C^2(\bar{\Omega})$ at some point $a \in \Omega \subset \mathbb{R}^2$. For this, we use the regularized output functional

$$J_\varepsilon(u) := |B_\varepsilon(a)|^{-1} \int_{B_\varepsilon(a)} \partial_1 u(a)\, dx = \partial_1 u(a) + \mathcal{O}(\varepsilon^2), \quad \varepsilon := TOL.$$

The corresponding dual solution behaves like a regularized 'derivative Green function' of the Laplacian:

$$|\nabla^k z(x)| \approx r(x)^{-k-1} := (|x-a|^2 + \varepsilon^2)^{-(k+1)/2}.$$

Then, for bilinear elements, the a posteriori error estimate takes the form

$$|J_\varepsilon(e)| \approx \eta_\omega := \sum_{K \in \mathbb{T}_h} \frac{h_K^3}{r_K^3} \rho_K, \qquad r_K := \max_{x \in K} r(x).$$

Now, assume that the local residuals are related to the local mesh size like

$$\rho_K \approx h_K, \tag{5.1}$$

uniformly for every $K \in \mathbb{T}_h$ and $h > 0$. This condition may be checked a posteriori in the course of the mesh adaptation process. Then, we obtain

$$\eta_\omega \approx \sum_{K \in \mathbb{T}_h} \frac{h_K^4}{r_K^3}. \tag{5.2}$$

In Section 4.2, we have seen that the optimal mesh for prescribed accuracy TOL is characterized by the equilibration property

$$\eta_K = \frac{h_K^4}{r_K^3} \approx \frac{TOL}{N} \quad \Rightarrow \quad |J_\varepsilon(e)| \approx \sum_{K \in \mathbb{T}_h} \frac{\mathrm{TOL}}{N} \approx \mathrm{TOL}.$$

From this, we derive

$$h_K^2 \approx r_K^{3/2} \Big(\frac{TOL}{N}\Big)^{1/2},$$

and consequently,

$$N = \sum_{K \in \mathbb{T}_h} h_K^2 h_K^{-2} = \Big(\frac{N}{TOL}\Big)^{1/2} \sum_{K \in \mathbb{T}_h} h_K^2 r_K^{-3/2} \approx \Big(\frac{N}{TOL}\Big)^{1/2}.$$

This implies that

$$N \propto TOL^{-1}, \tag{5.3}$$

which is better than the $N \propto TOL^{-2}$ that could be achieved on uniformly refined meshes on the basis of the general *a priori* convergence estimate

$$|\partial_1 e(a)| = \mathcal{O}(h).$$

This predicted asymptotic behavior is well confirmed by the results shown in Table 5.1. We emphasize that in this example strong mesh refinement occurs, although the solution is smooth. In fact, this phenomenon should rather be interpreted as 'mesh coarsening' away from the point of evaluation. Further, observing that $r_{\min} \sim TOL$ and $r_{\max} \sim 1$, for the optimized mesh, there holds

$$h_{\min} \approx TOL^{5/4}, \quad h_{\max} \approx TOL^{1/2},$$

and, consequently, $h_{\min} \approx h_{\max}^{5/2}$, which means that

$$L \approx \frac{5}{4} \frac{\log(TOL)}{\log(2)}$$

refinement cycles are needed to reach the 'optimal' mesh.

Table 5.1: *Computing* $\partial_1 u(0)$ *using the estimator* η_ω *(L refinement levels); from Becker and Rannacher [30].*

TOL	N	L	$\lvert J_\varepsilon(e)\rvert$	η_ω
4^{-3}	940	9	$4.10{\cdot}10^{-1}$	$1.42{\cdot}10^{-2}$
4^{-4}	4912	12	$4.14{\cdot}10^{-3}$	$3.50{\cdot}10^{-3}$
4^{-5}	20980	15	$2.27{\cdot}10^{-4}$	$9.25{\cdot}10^{-4}$
4^{-6}	86740	17	$5.82{\cdot}10^{-5}$	$2.38{\cdot}10^{-4}$

Remark 5.1. Relation (5.1) is decisive for the optimality of a refined mesh. To see this, suppose that only

$$\rho_K \approx h_K^{1-\varepsilon}$$

holds, for some small $\varepsilon > 0$. Then, the above calculation would result in

$$h_K^2 \approx r_K^{6/(4-\varepsilon)} \Big(\frac{TOL}{N}\Big)^{2/(4-\varepsilon)},$$

and consequently,

$$N = \sum_{K\in\mathbb{T}_h} h_K^2 h_K^{-2} = \Big(\frac{N}{TOL}\Big)^{2/(4-\varepsilon)} \sum_{K\in\mathbb{T}_h} h_K^2 r_K^{-6/(4-\varepsilon)} \propto \Big(\frac{N}{TOL}\Big)^{2/(4-\varepsilon)}.$$

This would give us the asymptotic complexity

$$N \propto TOL^{-1-\varepsilon/2},$$

which growths faster than TOL^{-1}.

Remark 5.2. An alternative, more explicit strategy for mesh adaptation may be based on the balancing condition

$$\frac{h_K \rho_K}{r_K^3} \propto \frac{\mathrm{TOL}}{|\Omega|} \quad \Rightarrow \quad |J_\varepsilon(e)| \propto \sum_{K\in\mathbb{T}_h} h_K^2 \frac{\mathrm{TOL}}{|\Omega|} \approx \mathrm{TOL}.$$

Here, the complexity analysis gives us (assuming again that $\rho_K \approx h_K$)

$$h_K \propto \Big(\frac{\mathrm{TOL}}{|\Omega|}\Big)^{1/2} r_K^{3/2},$$

and, consequently,

$$N = \sum_{K\in\mathbb{T}_h} h_K^2 h_K^{-2} \propto \frac{|\Omega|}{\mathrm{TOL}} \sum_{K\in\mathbb{T}_h} h_K^2 r_K^{-3} \propto \frac{|\Omega|}{\mathrm{TOL}^2},$$

which shows that this approach is not efficient for very singular error functionals.

5.1 Convergence of residuals

The above example illustrates that the assumed asymptotic relation (5.1) for the residuals ρ_K is crucial for deriving optimal complexity results. For analyzing this question in the case of d-linear finite elements, it suffices to consider the edge-residual part $\|[\partial_n u_h]\|_{\partial K}$, since on Cartesian meshes the cell-residual term automatically satisfies

$$\|f+\Delta u_h\|_K = \|f\|_K \leq c(f)h_K,$$

for bounded f. Now, notice that

$$h_K^{-3/2}\|[\partial_n u_h]\|_{\partial K} =: \left|D_h^2 u_{h|K}\right|$$

can be viewed as a mean value of a second-order difference quotient of u_h on the cell-patch $\tilde{K}$ containing K and its neighbors. Hence, in order to establish the bound (5.1), we have to seek for an estimate of the form

$$\sup_{h>0}\max_{x\in\Omega}|D_h^2 u_h| \leq c(u), \tag{5.4}$$

where the constant $c(u)$ depends on bounds for higher-order derivatives of the solution u. For proving (5.4), one may use the local inverse relation,

$$\|\nabla q\|_K \leq c_r h_K^{-1}\|q\|_K, \quad q \in P_r(K),$$

and the natural nodal interpolation $I_h u \in V_h$,

$$\begin{aligned}
|D_h^2 u_{h|K}| &\leq h_K^{-1}\|D_h^2 u_h\|_K \\
&\leq h_K^{-1}\|D_h^2(u_h - I_h u)\|_K + h_K^{-1}\|D_h^2 I_h u\|_K \\
&\leq ch_K^{-2}\|\nabla e\|_K + ch_K^{-2}\|\nabla(u-I_h u)\|_K + h_K^{-1}\|D_h^2 I_h u\|_K \\
&\leq ch_K^{-2}\|\nabla e\|_K + ch_K^{-1}\|\nabla^2 u\|_{\tilde{K}} \\
&\leq c\|\nabla^2 u\|_\infty,
\end{aligned}$$

where again $\tilde{K}$ denotes a cell-patch neighborhood of K. Unfortunately, this argument only works on *quasi-uniform* meshes, for which $h_{\max}/h_{\min} \leq c$ is assumed, since the local error estimate

$$\|\nabla e\|_{\infty;K} \leq h_K c(u)$$

does not hold in this strong form on meshes with $h_{\max}/h_{\min} \to \infty$. What one can prove is the weaker version

$$\|\nabla e\|_{\infty;K} \leq c(u)\Big\{\max_{K'\in\hat{S}(K)} h_{K'} + h^2\Big\},$$

where $h := h_{max}$, and $\hat{S}(K)$ is some $\mathcal{O}(1)$-neighborhood of the cell K. Alternatives may be found by adopting weighted-norm techniques from the 'classical' pointwise error analysis of finite element approximation. However, most of these results also require the mesh to be quasi-uniform. Hence, the problem must be left open.

5.2 Approximation of weights

Next, we want to analyze the effect of approximating the dual solution z in the weights ω_K, on the accuracy of the error estimate. For simplicity, we again restrict our attention to the two-dimensional case. In the following, we will examine two different methods of approximating z:

i) Approximation by a higher-order method. First, we consider the approximation of the dual solution z by its Ritz projection $z_h^{(2)}$ into the space $V_h^{(2)}$ of biquadratic finite elements, defined by

$$(\nabla\varphi_h, \nabla z_h^{(2)}) = J(\varphi_h) \quad \varphi_h \in V_h^{(2)}.$$

The resulting approximate error representation then reads

$$\tilde{E}(u_h) := \sum_{K\in\mathbb{T}_h} \{(R_h, z_h^{(2)} - I_h z_h^{(2)})_K + (r_h, z_h^{(2)} - I_h z_h^{(2)})_{\partial K}\}.$$

Its difference to the exact error representation $E(u_h)$ can be written in the form

$$E(u_h) - \tilde{E}(u_h) = \rho(u_h)(\tilde{e}^* - I_h\tilde{e}^*) = (\nabla e, \nabla(\tilde{e}^* - I_h\tilde{e}^*)),$$

with the abbreviation $\tilde{e}^* := z - z_h^{(2)}$. For estimating this error, we recall the well-known a priori error estimate for biquadratic finite elements:

$$\Big(\sum_{K\in\mathbb{T}_h} \|\nabla^k\tilde{e}^*\|_K^2\Big)^{1/2} \le ch^{3-k}\|\nabla^3 z\|, \quad k = 0, 1, 2.$$

Using this, we can then estimate as follows:

$$\begin{aligned} |E(u_h) - \tilde{E}(u_h)| &\le \|\nabla e\|\,\|\nabla(\tilde{e}^* - I_h\tilde{e}^*\| \\ &\le ch\|\nabla^2 u\|\Big(\sum_{K\in\mathbb{T}_h} h_K^2\|\nabla^2\tilde{e}^*\|_K^2\Big)^{1/2} \\ &\le ch^3\|\nabla^2 u\|\,\|\nabla^3 z\|. \end{aligned} \tag{5.5}$$

This estimate is unsatisfactory, as it requires the primal as well as the dual solution to be smooth which rules out most interesting applications. However, this objection can be somewhat weakened by the following modifications of the argument:

$$\begin{aligned} |E(u_h) - \tilde{E}(u_h)| &= |(\nabla e, \nabla\tilde{e}^*)| = |(\nabla(u - I_h u), \nabla\tilde{e}^*)| \\ &\le ch^2\Big(\sum_{K\in\mathbb{T}_h} h_K^2\|\nabla^2 u\|_K^2\Big)^{1/2}\|\nabla^3 z\|, \end{aligned} \tag{5.6}$$

or, setting $\tilde{e} := u - u_h^{(2)}$ the 'biquadratic' Ritz projection error,

$$|E(u_h) - \tilde{E}(u_h)| = |(\nabla e, \nabla \tilde{e}^*)| = |(\nabla \tilde{e}, \nabla(z - I_h z))|$$
$$\leq ch^2 \|\nabla^3 u\| \Big(\sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla^2 z\|_K^2 \Big)^{1/2}. \tag{5.7}$$

Here only smoothness of one of the solutions u and z is required and singularities in the other one can be compensated by proper mesh refinement. Though the estimates (5.6) and (5.7) are better than (5.5), they still do not cover point-error evaluation, i.e., $J(u) = u(x_0)$, since in this case

$$\sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla^2 z\|_K^2 \approx \sum_{K \in \mathbb{T}_h} h_K^2 r_K^{-4} \approx \mathcal{O}(1),$$

and generally h^2 is not smaller than TOL. To get a better result, we may use a L^∞-L^1-duality argument as follows:

$$|E(u_h) - \tilde{E}(u_h)| = |(\nabla e, \nabla \tilde{e}^*)| = |(\nabla(u - I_h^{(2)} u), \nabla \tilde{e}^*))|$$
$$\leq c \max_K \{h_K^2 \|\nabla^3 u\|_{\infty;K}\} \int_\Omega |\nabla \tilde{e}^*| \, dx \tag{5.8}$$
$$\leq ch |\log(h_{min})| \max_K \{h_K^2 \|\nabla^3 u\|_{\infty;K}\}.$$

The L^1-error estimate for the regularized Green function,

$$\int_\Omega |\nabla \tilde{e}^*| \, dx \leq c |\log(h_{min})|,$$

used in deriving (5.8) has been proven in Frehse and Rannacher [60]) only for quasi-uniform meshes; however, its extension to locally refined meshes is a technical exercise. The estimates (5.5)–(5.8) are useful provided that $h^3 \ll TOL$ on the current mesh. According to the discussion at the beginning, this is satisfied even in the case of derivative point value evaluation with

$$TOL \approx h_{max}^2.$$

However, since computing the dual solution by higher-order elements is too expensive in most practical situations, we will not pursue this discussion further.

ii) Approximation by higher-order interpolation. Next, we consider the theoretical justification of the approximate error estimator

$$\tilde{E}(u_h) := \sum_{K \in \mathbb{T}_h} \{(R_h, I_{2h}^{(2)} z_h - z_h)_K + (r_h, I_{2h}^{(2)} z_h - z_h)_{\partial K}\},$$

where $I_{2h}^{(2)} z_h$ is the patch-wise *biquadratic* interpolation of the *bilinear* Ritz projection z_h as defined above. This raises the question: *Why should* $I_{2h}^{(2)} z_h$ *be a better approximation to* z *than* z_h*?* In fact, the construction of $I_{2h}^{(2)} z_h$ is based on nodal point information of z_h, and the point error $(z-z_h)(a)$ behaves generally not better then $\mathcal{O}(h^2)$, even on uniform meshes. Hence, it seems unlikely that

$$\|z-I_{2h}^{(2)} z_h\|_K \ll \|z-z_h\|_K.$$

However, this may not be the right point of view. We could also seek to prove the somewhat weaker relation

$$|\rho(u_h)(z-I_{2h}^{(2)} z_h)| \ll |\rho(u_h)(z-z_h)|,$$

which has the flavor of a global 'super-approximation' property. Therefore, it will probably depend on some uniformity property of the mesh. In order to pursue this thought further, we rewrite the error identity $J(e) = \rho(u_h)(z-z_h)$ in the form

$$J(e) = \rho(u_h)(z-I_{2h}^{(2)} z) + \rho(u_h)(I_{2h}^{(2)} z-I_{2h}^{(2)} z_h) + \rho(u_h)(I_{2h}^{(2)} z_h-z_h), \qquad (5.9)$$

where the last term on the right is just our approximate error estimator. The first and second term will be estimated separately. To this end, we have to assume that the meshes have been optimized such that

$$\rho_K \le c h_K, \quad K \in \mathbb{T}_h. \qquad (5.10)$$

Using the local approximation properties of the interpolation $I_{2h}^{(2)} z$,

$$\left(\|z-I_{2h}^{(2)} z\|_K^2 + \tfrac{1}{2} h_K \|z-I_{2h}^{(2)} z\|_{\partial K}^2\right)^{1/2} \le c_I^{(2)} h_K^3 \|\nabla^3 z\|_{S(K)},$$

where $S(K)$ denotes the cell-patch on which $I_{2h}^{(2)} z$ is defined, we obtain for the first term in (5.9):

$$|\rho(u_h)(z-I_{2h}^{(2)} z)| \le c\Big(\sum_{K\in\mathbb{T}_h} h_K^6 \rho_K^2\Big)^{1/2} \|\nabla^3 z\|.$$

Consequently, using (5.10) we arrive at the estimate

$$|\rho(u_h)(z-I_{2h}^{(2)} z)| \le c(u,z) h^3. \qquad (5.11)$$

The second term in (5.9) is the 'hard' one which requires more work, as it relates properties of the non-local Ritz projection and the local interpolation. Its estimation strongly relies on uniformity properties of the mesh $\mathbb{T}_h$. The idea is that the scaled error $h^{-2}e$ is a 'smooth' function, such that it can be approximated in V_h. To make this concept clear, suppose that the mesh $\mathbb{T}_h$ is *uniform* with mesh-width h. Then, it is known that in the nodal points, the error $z-z_h$ allows

an asymptotic expansion in powers of h which can be expressed in the form (see Blum and Rannacher [34])

$$I_h z - z_h = I_h(z - z_h) = h^2 I_h w + h^3 \tau_h,$$

with some h-independent function $w \in H_0^1(\Omega)$ and a remainder satisfying $\|\tau_h\| \leq c\|\nabla^3 z\|$. From this, noting that $I_{2h}^{(2)} z = I_{2h}^{(2)} I_h z$, we conclude

$$\rho(u_h)(I_{2h}^{(2)} z - I_{2h}^{(2)} z_h) = h^2 \rho(u_h)(I_{2h}^{(2)} w) + h^3 \rho(u_h)(\tau_h),$$

and, using Galerkin orthogonality,

$$\rho(u_h)(I_{2h}^{(2)} z - I_{2h}^{(2)} z_h) = h^2 \rho(u_h)(I_{2h}^{(2)} w - I_h w) + h^3 \rho(u_h)(\tau_h).$$

Assuming that the interpolation operators I_h and $I_{2h}^{(2)}$ behave like the H^1-stable Clément operator, we have

$$\|I_{2h}^{(2)} w - I_h w\|_K + \tfrac{1}{2} h^{1/2} \|I_{2h}^{(2)} w - I_h w\|_{\partial K} \leq ch \|\nabla w\|_{S_K}.$$

Collecting the above estimates, and using again (5.10), we find

$$|\rho(u_h)(I_{2h}^{(2)} z - I_{2h}^{(2)} z_h)| \leq c(u, z) h^3. \tag{5.12}$$

Finally, inserting the estimates (5.11) and (5.12) into (5.9), we conclude the desired estimate

$$J(e) = \tilde{E}(u_h) + \mathcal{O}(h^3). \tag{5.13}$$

We emphasize that this estimate has been derived for a smooth dual solution z which excludes almost all interesting applications. Furthermore, the meshes are required to be uniform which conflicts with our ultimate goal of mesh adaptation. Therefore, a really meaningful analysis has to deal with the following three complications:

- The estimates are to be proven on locally refined meshes with only limited uniformity properties.

- The estimates are to be localized in order to allow for singularities in the dual solution.

- Cases of non-smooth solutions u, either due to data irregularities or due to reentrant corners (the more severe case) should be considered.

with a cell patch $\tilde{K}$ around K, if u is not regular (e.g., corner singularities) with constants independent of h. The first relation can be proven for quasi-uniform meshes, but is still open for general locally refined meshes including hanging nodes. We want to check these conditions by numerical experiments.

a) Consider the Poisson model problem

$$-\Delta u = f \text{ in } \Omega, \quad u_{|\partial\Omega} = 0,$$

on the domain $\Omega = (-1,1)^2$ with $f(x) = \frac{1}{2}\pi^2 \cos(\frac{1}{2}x_1\pi)\cos(\frac{1}{2}x_2\pi)$. Apply the duality-based mesh refinement strategy for computing the two different functional values:

$$J(u) = \partial_1 u(0.5, 0.5), \qquad J(u) = \int_{-1}^{1} \partial_1 u(1, x_2)\, dx_2,$$

and monitor the behavior of $D_h^2 u_{h|K}$ for increasingly refined meshes.

b) Consider the Poisson model problem

$$-\Delta u = f \text{ in } \Omega, \quad u_{|\partial\Omega} = 0,$$

on the slit-domain $\Omega = (-1,1)^2 \setminus \{x \in \mathbb{R}^2, x_1 = 0, -1 \leq x_2 \leq 0\}$ again with $f(x) = \frac{1}{2}\pi^2 \cos(\frac{1}{2}x_1\pi)\cos(\frac{1}{2}x_2\pi)$. Apply again the duality-based mesh refinement strategy to compute the same functional values as above, and monitor the behavior of the quantity

$$\|\tilde{D}_h^2 u_h\|_\infty := \max_{K \in \mathbb{T}_h} \{r_K^{3/2} h_K^{-3/2} \|[\partial_n u_h]\|_K, \},$$

which reflects the expected singular behavior of u at the tip of the slit, where r_K is the distance of the cell K from this point. Interpret the observed results.

Chapter 6

An Abstract Approach for Nonlinear Problems

In this chapter, we will present a very general approach to a posteriori error estimation for the Galerkin approximation of nonlinear variational problems as developed in Becker and Rannacher [31]. The framework is kept on an abstract level in order to allow later for a unified application to rather different situations, such as nonlinear PDEs, but also eigenvalue and optimization problems. We prepare for this by recalling the special situation of a linear Galerkin approximation of the form

$$a(u,\psi) = F(\psi) \qquad \forall \psi \in V, \tag{6.1}$$

$$a(u_h,\psi_h) = F(\psi_h) \quad \forall \psi_h \in V_h, \tag{6.2}$$

with a linear functional $F(\cdot)$ as right-hand side. For a given (linear) output functional $J(\cdot)$ the associated continuous and discrete dual problems are

$$a(\psi,z) = J(\psi) \qquad \forall \psi \in V, \tag{6.3}$$

$$a(\psi_h,z_h) = J(\psi_h) \quad \forall \psi_h \in V_h. \tag{6.4}$$

Then, using Galerkin orthogonality, for the 'primal' error $e := u-u_h$ and the 'dual' error $e^* := z-z_h$, there holds

$$J(e) = a(e,z) = a(e,e^*) = a(u,e^*) = F(e^*),$$

and, introducing the weighted *primal* and *dual* residuals,

$$J(e) = F(z-\psi_h) - a(u_h,z-\psi_h) =: \rho(u_h)(z-\psi_h), \qquad \psi_h \in V_h, \tag{6.5}$$

$$F(e^*) = J((u-\varphi_h) - a(u-\varphi_h,z_h) =: \rho^*(z_h)(u-\varphi_h), \quad \varphi_h \in V_h. \tag{6.6}$$

This trivially implies the 'linear' error representation

$$J(e) = \tfrac{1}{2}\rho(u_h)(z-\psi_h) + \tfrac{1}{2}\rho^*(z_h)(u-\varphi_h), \quad \varphi_h,\psi_h \in V_h. \tag{6.7}$$

Below, we will see that this identity has a natural generalization to nonlinear problems, where the weighted primal and dual residuals are not identical.

6.1 Galerkin approximation of nonlinear equations

The following theory will be presented within an abstract functional analytic framework making as little assumptions as possible. Let $A(u)(\cdot)$ be a semilinear form and $J(\cdot)$ an output functional, not necessarily linear, defined on some function space V. The goal is the evaluation of $J(u)$ from the solution of the variational problem

$$A(u)(\psi) = 0 \quad \forall \psi \in V. \tag{6.8}$$

The corresponding Galerkin approximation uses finite dimensional subspaces $V_h \subset V$ to determine $u_h \in V_h$ by

$$A(u_h)(\psi_h) = 0 \quad \forall \psi_h \in V_h. \tag{6.9}$$

We assume the existence of directional derivatives of A and J up to order three denoted by $A'(u)(\varphi,\cdot)$, $A''(u)(\psi,\varphi,\cdot)$, $A'''(u)(\xi,\psi,\varphi,\cdot)$, and $J'(u)(\varphi)$, $J''(u)(\psi,\varphi)$, $J'''(u)(\xi,\psi,\varphi)$, respectively, for increments $\varphi,\psi,\xi \in V$. In these forms the dependence on the first argument in parentheses may be nonlinear while the dependence on all further arguments in the second set of parentheses is linear.

Example 6.1. A typical example of a nonlinear problem of the form we are interested in is the so-called *vector Burgers equation*

$$-\nu\Delta u + u\cdot\nabla u = f, \ \text{ in } \Omega, \quad u_{|\partial\Omega} = 0.$$

for a vector function $u \in H_0^1(\Omega)^d$. This is the natural generalization of the classical 1-D Burgers equation to multiple dimensions. The corresponding variational formulation has the form (6.8) with the semilinear form

$$A(u)(\psi) := \nu(\nabla u, \nabla\psi) + (u\cdot\nabla u, \psi) - (f,\psi).$$

Example 6.2. Another example of a nonlinear problem which will be considered in several exercises below is the diffusion-reaction equation

$$-\Delta u - u^3 = f, \ \text{ in } \Omega, \quad u_{|\partial\Omega} = 0.$$

for a scalar function $u \in H_0^1(\Omega)$. This problem is interesting in the context of bifurcation theory. In this case the corresponding semilinear form is given by

$$A(u)(\psi) := (\nabla u, \nabla\psi) - (u^3,\psi) - (f,\psi).$$

For estimating the error $J(u)-J(u_h)$, we employ the Euler-Lagrange method of constrained optimization. Introducing a 'dual' variable $z \in V$ ('adjoint' variable or 'Lagrangian multiplier'), we define the Lagrangian functional

$$\mathcal{L}(u,z) := J(u) - A(u)(z),$$

and seek for stationary points $\{u,z\} \in V \times V$ of $\mathcal{L}(\cdot,\cdot)$, i.e.,

$$\mathcal{L}'(u,z)(\varphi,\psi) = \left\{ \begin{array}{l} J'(u)(\varphi) - A'(u)(\varphi,z) \\ -A(u)(\psi) \end{array} \right\} = 0 \quad \forall \{\varphi,\psi\}. \tag{6.10}$$

Clearly, the u-component of any such stationary point is a solution of the original problem (6.8). The Galerkin approximations $\{u_h,z_h\} \in V_h \times V_h$ are defined by the discrete Euler-Lagrange system

$$\mathcal{L}'(u_h,z_h)(\varphi_h,\psi_h) = \left\{ \begin{array}{l} J'(u_h)(\varphi_h) - A'(u_h)(\varphi_h,z_h)\rangle \\ -A(u_h)(\psi_h) \end{array} \right\} = 0 \quad \forall \{\varphi_h,\psi_h\}, \tag{6.11}$$

where, again, the u_h-component of any stationary point is a solution of the discrete problem (6.9). Now, the goal is to estimate the error $J(u)-J(u_h)$ in terms of the residuals associated with this set of equations. We prepare for this by considering first the general situation of the Galerkin approximation of stationary points of functionals.

Proposition 6.1. *Let $L(\cdot)$ be a three-times differentiable functional defined on a (real or complex) vector space X which has a stationary point $x \in X$, i.e.,*

$$L'(x)(y) = 0 \quad \forall y \in X. \tag{6.12}$$

Suppose that on a finite dimensional subspace $X_h \subset X$, the corresponding Galerkin approximation

$$L'(x_h)(y_h) = 0 \quad \forall y_h \in X_h, \tag{6.13}$$

has a solution $x_h \in X_h$. Then, there holds the error representation

$$L(x) - L(x_h) = \tfrac{1}{2} L'(x_h)(x-y_h) + \mathcal{R}_h, \quad y_h \in X_h, \tag{6.14}$$

with a remainder term $\mathcal{R}_h$ which is cubic in the error $e^x := x-x_h$,

$$\mathcal{R}_h := \tfrac{1}{2} \int_0^1 L'''(x_h+se^x)(e^x,e^x,e^x)\, s(s-1)\, ds.$$

Proof. Since $L'(x)(e^x) = 0$, we can write

$$\begin{aligned} L(x) - L(x_h) = \int_0^1 & L'(x_h+se^x)(e^x)\, ds \\ & + \tfrac{1}{2} L'(x_h)(e^x) - \tfrac{1}{2}\{L'(x_h)(e^x) + L'(x)(e^x)\}. \end{aligned}$$

The last term on the right is just the approximation of the integral term by the trapezoidal rule. For this, we have the well-known error representation

$$\int_0^1 f(t)\,dt = \tfrac{1}{2}\{f(0)+f(1)\} + \tfrac{1}{2}\int_0^1 f''(s)s(s-1)\,ds.$$

Hence, we obtain

$$L(x) - L(x_h) = \tfrac{1}{2}L'(x_h)(e^x) + \tfrac{1}{2}\int_0^1 L'''(x_h+se^x)(e^x,e^x,e^x)\,s(s-1)\,ds.$$

Finally, observing that $L'(x_h)(y_h) = 0$ for all $y_h \in X_h$, we have that

$$L'(x_h)(e^x) = L'(x_h)(x-y_h), \quad y_h \in X_h.$$

This completes the proof. □

As an immediate consequence of Proposition 6.1, we obtain the following result for the Galerkin approximation of variational equations.

Proposition 6.2. *For any solution of equations (6.8) and (6.9), we have the error representation*

$$J(u) - J(u_h) = \tfrac{1}{2}\rho(u_h)(z-\psi_h) + \tfrac{1}{2}\rho^*(u_h,z_h)(u-\varphi_h) + \mathcal{R}_h^{(3)}, \tag{6.15}$$

with arbitrary $\varphi_h, \psi_h \in V_h$, and the 'primal' and 'dual' residuals

$$\rho(u_h)(\cdot) := -A(u_h)(\cdot),$$
$$\rho^*(u_h,z_h)(\cdot) := J'(u_h)(\cdot) - A'(u_h)(\cdot,z_h).$$

The remainder term $\mathcal{R}_h^{(3)}$ is cubic in the 'primal' and 'dual' errors $e := u-u_h$ and $e^* := z-z_h$,

$$\mathcal{R}_h^{(3)} = \tfrac{1}{2}\int_0^1 \{J'''(u_h+se)(e,e,e) - A'''(u_h+se)(e,e,e,z_h+se^*) - 3A''(u_h+se)(e,e,e^*)\}\,s(s-1)\,ds.$$

Proof. In order to apply Proposition 6.1, we define the space $X := V\times V$ and for arguments $x = \{u,z\} \in X$ the functional $L(x) := \mathcal{L}(u,z)$. In this context the stationary points are denoted by $x := \{u,z\}$ and $x_h := \{u_h,z_h\}$ with the error $e^x := x-x_h$. Then, we have

$$J(u) - J(u_h) = L(x) - A(u)(z) - L(x_h) + A_h(u_h)(z_h) = L(x) - L(x_h).$$

Hence, the error representation of Proposition 6.1 gives us

$$J(u) - J(u_h) = \tfrac{1}{2}L'(x_h)(x-y_h) + \mathcal{R}_h, \quad y_h \in X_h,$$

with

$$\mathcal{R}_h := \tfrac{1}{2}\int_0^1 L'''(x_h+se^x)(e^x,e^x,e^x)\,s(s-1)\,ds.$$

By construction, we have for arbitrary $y_h = \{\varphi_h, \psi_h\} \in X_h$:

$$\begin{aligned} L'(x_h)(x-y_h) &= \mathcal{L}'_u(u_h,z_h)(u-\varphi_h) + \mathcal{L}'_z(u_h,z_h)(z-\psi_h) \\ &= J'(u_h)(u-\varphi_h) - A'(u_h)(u-\varphi_h,z_h) - A(u_h)(z-\psi_h) \\ &= \rho^*(u_h,z_h)(u-\varphi_h) + \rho(u_h)(z-\psi_h). \end{aligned}$$

Notice that $\mathcal{L}(u,z)$ is linear in z. Consequently, the third derivative of $L(\cdot)$ consists of only three terms, namely,

$$J'''(u_h+se)(e,e,e) - A'''(u_h+se)(e,e,e,z_h+se^*) - 3A''(u_h+se)(e,e,e^*).$$

This implies the asserted form of the remainder term $\mathcal{R}_h^{(3)}$. □

Remark 6.3. The derivation of the error representations (6.14) and (6.15) does not require the uniqueness of solutions; this is important, for example, for the application to eigenvalue problems. In cases with non-unique solutions, the a priori assumption $x_h \to x$ $(h \to 0)$ makes the result meaningful as then the remainder term can be assumed to be small.

Remark 6.4. The actual evaluation of the error identity requires guesses for the primal and dual solutions u and z which are usually computed from the Galerkin approximations u_h and z_h by some post-processing as described in Section 4.1.

Remark 6.5. The cubic remainder $\mathcal{R}_h^{(3)}$ can usually be neglected. However, in parameter-dependent problems when approaching a bifurcation point, the derivatives of $A(u)(\cdot)$ and consequently $\mathcal{R}_h^{(3)}$ may become large. In such a situation the abstract theory can still be applied but with special care. The extreme situation will be seen in Chapter 7 in the context of eigenvalue problems where one is directly working in the singular case.

In the linear case, we have seen that the primal and dual residual terms coincide, i.e. $\rho(u_h)(z-\psi_h) = \rho^*(z_h)(u-\varphi_h)$. This is no longer true in the nonlinear case, but the deviation from this property, i.e. the degree of nonlinearity of the problem, can be estimated as the following proposition shows.

Proposition 6.6. *With the notation from above, there holds*

$$\rho^*(u_h,z_h)(u-\varphi_h) = \rho(u_h)(z-\psi_h) + \Delta\rho, \tag{6.16}$$

for any $\varphi_h, \psi_h \in V_h$, *with*

$$\Delta\rho := \int_0^1 \left\{A''(u_h+se)(e,e,z_h+se^*) - J''(u_h+se)(e,e)\right\} ds.$$

Further, we have the simplified error representation

$$J(u) - J(u_h) = \rho(u_h)(z-\varphi_h) + \mathcal{R}_h^{(2)}, \tag{6.17}$$

for any $\varphi_h \in V_h$, *with the quadratic remainder*

$$\mathcal{R}_h^{(2)} := \int_0^1 \{A''(u_h+se)(e,e,z) - J''(u_h+se)(e,e)\}\, s\, ds$$

Proof. We introduce the scalar function $g(\cdot)$ by

$$g(s) := J'(u_h+se)(e) - A'(u_h+se)(e, z_h+se^*).$$

By the definition of z and z_h, there holds

$$\begin{aligned} g(1) &= J'(u)(e) - A'(u)(e,z) = 0, \\ g(0) &= J'(u_h)(e) - A'(u_h)(e,z_h) = \rho^*(u_h,z_h)(e), \end{aligned}$$

and

$$\begin{aligned} g'(s) = {} & J''(u_h+se)(e,e) - A''(u_h+se)(e,e,z_h+se^*) \\ & - A'(u_h+se)(e,e^*). \end{aligned}$$

Therefore, using Galerkin orthogonality,

$$\begin{aligned} \rho^*(u_h,z_h)(u-\varphi_h) = \rho^*(u_h,z_h)(e) = g(0) = g(0) - g(1) &= -\int_0^1 g'(s)\, ds \\ &= \int_0^1 \{A''(u_h+se)(e,e,z_h+se^*) - J''(u_h+se)(e,e)\}\, ds \\ &\quad + \int_0^1 A'(u_h+se)(e,e^*)\, ds \\ &= \Delta\rho + \rho(u_h)(e^*) = \Delta\rho + \rho(u_h)(z-\psi_h). \end{aligned}$$

this proves (6.16). In order to prove (6.17), we use integration by parts, obtaining

$$\begin{aligned} \mathcal{R}_h^{(2)} &= \int_0^1 \{A''(u_h+se)(e,e,z) - J''(u_h+se)(e,e)\}\, s\, ds \\ &= -\int_0^1 \{A'(u_h+se)(e,z) - J'(u_h+se)(e)\}\, ds + A'(u)(e,z) - J'(u)(e), \end{aligned}$$

where the last two terms vanish by definition of z. Consequently, employing again Galerkin orthogonality, we obtain

$$\mathcal{R}_h^{(2)} = -\rho(u_h)(z-\psi_h) + J(u) - J(u_h),$$

for arbitrary $\psi_h \in V_h$. This completes the proof.

We note that the simplified error representation (6.17) could have been derived also from (6.15) using the relation (6.16). However, this involves lengthy calculations, so that we preferred to present a more direct argument. □

In order to use the error representations (6.15) or (6.17) for practical mesh adaptation, we have to evaluate the primal and dual residual terms. As in the linear case, this requires approximation of the dual solution z and in the context of (6.15) additionally that of the primal solution u. This may be achieved again by post-processing of the Galerkin solutions z_h and u_h exploiting higher-order interpolation. Let the resulting approximations be denoted by $\tilde{z}_h$ and $\tilde{u}_h$. Then, neglecting the remainder terms the approximate error representations take the form

$$\tilde{E}(u_h, z_h) := \tfrac{1}{2}\rho(u_h)(\tilde{z}_h - z_h) + \tfrac{1}{2}\rho^*(u_h, z_h)(\tilde{u}_h - u_h), \tag{6.18}$$

and

$$\tilde{E}(u_h) := \rho(u_h)(\tilde{z}_h - z_h). \tag{6.19}$$

Remark 6.7. The identity (6.16) is useful as it offers the possibility of controlling the remainder $\mathcal{R}_h^{(2)}$ in the simplified error representation (6.17). In fact, comparing the two error representations (6.15), (6.17) and using (6.16), we see that

$$\begin{aligned}
\mathcal{R}_h^{(2)} &= -\rho(u_h)(z - \psi_h) + J(u) - J(u_h) \\
&= -\rho(u_h)(z - \psi_h) + \tfrac{1}{2}\rho(u_h)(z - \psi_h) + \tfrac{1}{2}\rho^*(u_h, z_h)(u - \varphi_h) + \mathcal{R}_h^{(3)} \\
&= \tfrac{1}{2}\rho^*(u_h, z_h)(u - \varphi_h) - \tfrac{1}{2}\rho(u_h)(z - \psi_h) + \mathcal{R}_h^{(3)} \\
&= \tfrac{1}{2}\Delta\rho + \mathcal{R}_h^{(3)}.
\end{aligned}$$

Hence, we may try to control the linearization error by a posteriori checking the condition

$$|\Delta\rho| \approx |\rho^*(u_h, z_h)(\tilde{u}_h - u_h) - \rho(u_h)(\tilde{z}_h - z_h)| \ll TOL. \tag{6.20}$$

where $\tilde{u}_h \approx u$ and $\tilde{z}_h \approx z$ are higher-order approximations and the cubic remainder term $\mathcal{R}_h^{(3)}$ is neglected.

Remark 6.8. The possibility of improving on the *linearization error*, i.e. reducing the remainder term $\mathcal{R}_h^{(2)}$, by post-processing the Ritz approximation u_h has been studied in Vexler [134]. This costly process may be relevant in cases when the problem to be solved is close to bifurcation.

Remark 6.9. A posteriori error estimates for the Galerkin finite element approximation of nonlinear variational problems have also been derived using extensions of the classical 'energy-norm-based' approach; see, e.g., Verfürth [131, 133]. These results rely on assumptions on the monotonicity of the underlying problem, i.e. the coercivity of certain derivative forms, or involve nonlinear stability constants which depend on the unknown solution. These estimates usually represent the 'worst case' scenario and will only in special cases be of practical value.

6.2 A nested solution approach

For solving the nonlinear problems by a Galerkin finite element method, we employ the following iterative scheme. Starting from a coarse initial mesh $\mathbb{T}_0$, a hierarchy of refined meshes

$$\mathbb{T}_0 \subset \mathbb{T}_1 \subset \cdots \subset \mathbb{T}_l \subset \cdots \subset \mathbb{T}_L,$$

and corresponding finite element spaces V_l, $l = 0, ..., L$, with dimensions N_l is generated by the following nested solution process:

1. *Initialization:* For $l = 0$, compute a solution $u_0 \in V_0$ on the mesh $\mathbb{T}_0$.

2. *Defect correction iteration:* For $l \geq 1$, start with $u_l^{(0)} = u_{l-1} \in V_l$. For a computed iterate $u_l^{(j)} \in V_l$ evaluate the defect

$$(d_l^{(j)}, \psi_l) = -A(u_l^{(j)})(\psi_l), \quad \forall \psi_l \in V_l,$$

and solve the correction equation (Newton update)

$$\tilde{A}'(u_l^{(j)})(v_l^{(j)}, \psi_l) = (d_l^{(j)}, \psi_l) \quad \forall \psi_l \in V_l,$$

by Krylov-space or multigrid iterations using the hierarchy of previously constructed meshes $\{\mathbb{T}_{l-1}, ..., \mathbb{T}_0\}$. Update $u_l^{(j+1)} = u_l^{(j)} + \alpha_l^{(j)} v_l^{(j)}$, with a step-length parameter $\alpha_l^{(j)}$, set $j = j+1$, and repeat the iteration. This process is carried until a limit $u_l \in V_l$, is reached with some required accuracy.

3. *Error estimation:* Solve the (linearized) discrete dual problem

$$z_l \in V_l: \quad A'(u_l)(\varphi_l, z_l) = J(\varphi_l) \quad \forall \varphi_l \in V_l,$$

and evaluate the a posteriori error estimate (6.18):

$$J(e_l) \approx \tilde{E}(u_l, z_l).$$

If $|\tilde{E}(u_l, z_l)| \leq TOL$, or $N_l \geq N_{\max}$, then stop. Otherwise cell-wise mesh adaptation yields the new mesh $\mathbb{T}_{l+1}$. Then, set $l := l + 1$ and go to (2).

Remark 6.10. The nonlinear iteration described above is oriented at the solution of stationary elliptic problems in which a global transfer of information is present. In the case of transport-dominated problems, particularly those with information transfer into one direction only such as in nonstationary problems, one would organize the solution process differently, taking this transport direction into account.

Remark 6.11. In the described Newton-like iteration the mesh adaptation is done for the limit solution on the current mesh in order to have a rigorous theoretical basis. However, it may be inefficient to carry the iteration on a coarser mesh to

the limit knowing that the discretization accuracy on this mesh is still insufficient. Hence, one would like to combine the estimation of the discretization error with that of the iteration error, both in accordance with the accuracy in the target quantity. Such a combined error estimator has been developed for a linear multigrid iteration in Becker et al. [27] and Becker [19], but it is an open questions for the nonlinear Newton iteration.

Remark 6.12. The solution of the *linear* dual problem usually requires much less work compared to solving the nonlinear primal problem (6.8). In fact, in the context of the nested solution method described above the primal solution is obtained by a Newton iteration which requires the solution of several linear problems until convergence is reached. Then, solving the linear dual problem for the converged primal solution normally corresponds to about one additional Newton step. Further, this extra work is spent on optimized meshes adapted to the particular goal of the computation. Hence, particularly for nonlinear problems, the duality-based approach to adaptivity becomes relatively 'cheap'. This is demonstrated by the examples from structural and fluid mechanics which will be presented in Chapter 10 and Chapter 11.

Remark 6.13. We remark on the following particular aspect in the evaluation of the dual residual $\rho^*(u_h, z_h)(u-\varphi_h)$. The discrete dual solution z_h is determined by the variational equation (6.11) which is the Galerkin discretization of the corresponding continuous variational problem (6.10). The latter can usually be interpreted as the 'weak' form of a certain system of differential equations. The cell and edge residuals R_h^* and r_h^* of the dual solution z_h are then taken with respect to the 'strong' form of this dual problem. However, in deriving these residuals, we have to observe their consistency with the variational formulation (6.11) in order not to miss terms which may not be present on the continuous level, such as for example the residual terms $\nabla \cdot v_{h|K}$ occurring in solving the incompressible Navier-Stokes equations. We will comment on this point later on when a 'naive' derivation of the dual cell and edge residuals may lead to wrong results (see Remark 11.2).

6.3 Exercises

Exercise 6.1. Consider the Galerkin approximation of the stationary 1-dimensional Burgers equation

$$-\nu u_{xx} + u u_x = f, \quad \text{in } (0,1), \quad u(0) = 0 = u(1),$$

by using piecewise linear finite elements. Determine the cubic and quadratic remainder terms in the a posteriori error representations for a *linear* output functional.

Exercise 6.2. Use the second-order 'nonlinear' a posteriori error estimate to derive the L^2-norm a posteriori error bound for the Galerkin finite element approximation of the (linear) Poisson problem.

Exercise 6.3 *(Practical exercise).* Consider the nonlinear diffusion-reaction equation

$$-\Delta u - u^3 = f, \text{ in } \Omega, \quad u_{|\partial\Omega} = 0,$$

where Ω is the domain defined in Exercise 3.3. The discretization is by the usual bilinear finite elements. For $f \equiv \alpha = 1, \ldots, 75$, compute the solution's mean value over the subdomain $\Omega' = \{x \in \Omega, x_1 > .5, x_2 > .5\} \subset \Omega$,

$$J_1(u) := 4 \int_{\Omega'} u(x)\, dx.$$

Monitor the size (relative to the estimated error) of the control term for the linearization error,

$$\Delta\tilde{\rho} := \rho^*(u_h, z_h)(\tilde{I}_{2h}^{(2)} u_h - u_h) - \rho(u_h)(\tilde{I}_{2h}^{(2)} z_h - z_h),$$

by approximating the weights using the *biquadratic patch-wise interpolations* $\tilde{I}_{2h}^{(2)} u_h$ and $\tilde{I}_{2h}^{(2)} z_h$ of the discrete primal and dual solutions u_h and z_h, respectively. Repeat the same calculations using the true *biquadratic Ritz projection* $z_h^{(2)}$ of z instead. Explain the observed results.

Chapter 7

Eigenvalue Problems

In the following, we will apply the abstract theory of the DWR method developed in Chapter 6 to error control in the approximation of eigenvalue problems. We mention some prototypical examples we are particularly interested in:

- The *symmetric* eigenvalue problem of the Laplace operator:

$$-\Delta u = \lambda u.$$

- The *nonsymmetric* eigenvalue problem of a convection-diffusion operator:

$$-\Delta u + b\cdot\nabla u = \lambda u.$$

- The *stability* eigenvalue problem governed by the linearized Navier-Stokes equations:

$$-\nu\Delta v + \hat{v}\cdot\nabla v + v\cdot\nabla\hat{v} + \nabla p = \lambda v, \quad \nabla\cdot v = 0,$$

  where $\hat{v}$ is some 'base solution' the stability of which is to be investigated.

The results we will present for the approximation of these problems are taken from Heuveline and Rannacher [77, 78].

The above eigenvalue problems are all treated within an abstract setting, which will be laid out in the following. Let V be a (complex) Hilbert space. We seek $\{u, \lambda\} \in V \times \mathbb{C}$ satisfying

$$\mathcal{A}u = \lambda \mathcal{M}u,$$

with linear operators $\mathcal{A}$ and $\mathcal{M}$ in V. The operator $\mathcal{M}$ is assumed to be self-adjoint and positive semi-definite. It is introduced to allow for the situation of the stability eigenvalue problem for the Navier-Stokes equations, where the eigenvalue term only occurs in the first of the two equations (the momentum equation). We do not go deeper into the abstract functional analytic setting of eigenvalue problems

since we will concentrate below on concrete problems formulated in the standard function spaces. Again, we prefer the variational formulation

$$a(u,\psi) = \lambda\, m(u,\psi) \quad \forall \psi \in V, \tag{7.1}$$

where $a(\cdot,\cdot) := \langle \mathcal{A}\cdot,\cdot\rangle$ is the (generally nonsymmetric) sesquilinear form generated by $\mathcal{A}$ and $m(\cdot,\cdot)$ is the symmetric semi-definite sesquilinear form corresponding to $\mathcal{M}$ (usually the L^2 scalar product). Eigenfunctions are assumed to be normalized by $m(u,u) = 1$. For nonsymmetric $\mathcal{A}$, one also considers the associated *adjoint eigenvalue problem* for the Hilbert-space adjoint $\mathcal{A}^*$ of $\mathcal{A}$,

$$\mathcal{A}^* u^* = \lambda^*\, \mathcal{M} u^*.$$

Observing that $a^*(u^*,\varphi) = \langle \mathcal{A}^* u^*, \varphi\rangle = \overline{a(\varphi,u^*)}$ and $m(u^*,\varphi) = \overline{m(\varphi,u^*)}$, this has the variational form

$$a(\varphi,u^*) = \bar\lambda^*\, m(\varphi,u^*) \quad \forall \varphi \in V. \tag{7.2}$$

Note that in the context of matrix eigenvalue problems the *dual eigenvectors* u^* are also called *left eigenvectors*. From the definition, we see that primal and dual eigenvalues are related by $\lambda^* = \bar\lambda$, while the corresponding eigenvectors may differ. Usually, the dual eigenvectors are normalized by requiring

$$m(u,u^*) = 1.$$

This normalization is possible only if u^* is not m-orthogonal with respect to the whole eigenspace of λ, which is equivalent to requiring that λ has trivial *defect*, i.e., its algebraic eigenspace reduces to the geometric one (only trivial Jordan blocks in the matrix case); for a more detailed discussion of this issue see Heuveline and Rannacher [77], and the literature cited therein. We will see that the simultaneous consideration of primal and dual eigenvalue problems is essential for rigorous a posteriori error estimation.

The Galerkin approximation of (7.1) and (7.2) uses finite dimensional subspaces $V_h \subset V$, as described above, to determine $\{u_h,\lambda_h\}, \{u_h^*,\lambda_h^*\} \in V_h \times \mathbb{C}$, satisfying

$$a(u_h,\psi_h) = \lambda_h\, m(u_h,\psi_h) \quad \forall \psi_h \in V_h, \quad m(u_h,u_h) = 1, \tag{7.3}$$

$$a(\varphi_h,u_h^*) = \bar\lambda_h^*\, m(\varphi_h,u_h^*) \quad \forall \varphi_h \in V_h, \quad m(u_h,u_h^*) = 1. \tag{7.4}$$

Our goal is to control the errors $\lambda - \lambda_h$, $u - u_h$, and $u^* - u_h^*$ in the eigenvalues and eigenfunctions in terms of the residuals associated with these equations.

7.1 A posteriori error analysis

In order to derive a posteriori error estimates, we embed the present situation into the general framework of variational equations as considered in Chapter 6.

Consider the product spaces $\mathcal{V} := V \times \mathbb{C}$ and $\mathcal{V}_h := V_h \times \mathbb{C} \subset \mathcal{V}$ and define for pairs $U := \{u, \lambda\} \in \mathcal{V}$ and $\Psi = \{\psi, \nu\} \in \mathcal{V}$ the semilinear form

$$A(U)(\Psi) := \lambda\, m(u, \psi) - a(u, \psi) + \overline{\nu}\left\{m(u, u) - 1\right\}.$$

With this notation, the above continuous eigenvalue problem and its discrete analogue can be written in the following compact form of semilinear variational equations:

$$A(U)(\Psi) = 0 \qquad \forall \Psi \in \mathcal{V}, \tag{7.5}$$

$$A(U_h)(\Psi_h) = 0 \qquad \forall \Psi_h \in \mathcal{V}_h, \tag{7.6}$$

where $U := \{u, \lambda\}$ and $U_h := \{u_h, \lambda_h\}$, respectively. In order to control the error in the approximation of the eigenvalues, we use the output functional

$$J(\Phi) := \mu\, m(\varphi, \varphi).$$

Since $m(u, u) = 1$ at the solution $U = \{u, \lambda\}$, there holds

$$J(U) = \lambda\, m(u, u) = \lambda,$$

i.e., as desired this functional picks out the eigenvalue. We recall the Lagrange approach from Chapter 6, particularly the Lagrangian functional for arguments $U = \{u, \lambda\}$ and $\Psi = \{\psi, \nu\}$:

$$\begin{aligned} \mathcal{L}(U, \Psi) &= J(U) - A(U)(\Psi) \\ &= \lambda\, m(u, u) - \lambda\, m(u, \psi) + a(u, \psi) - \overline{\nu}\left\{m(u, u) - 1\right\}. \end{aligned}$$

The dual solution $Z = \{z, \pi\} \in \mathcal{V}$ and its Galerkin approximation $Z_h = \{z_h, \pi_h\} \in \mathcal{V}_h$ are then determined by the equations

$$A'(U)(\Phi, Z) = J'(U)(\Phi) \qquad \forall \Phi \in \mathcal{V}, \tag{7.7}$$

$$A'(U_h)(\Phi_h, Z_h) = J'(U_h)(\Phi_h) \qquad \forall \Phi_h \in \mathcal{V}_h. \tag{7.8}$$

The left and right hand sides of (7.7) read, for $Z = \{z, \pi\}$, $U = \{u, \lambda\}$, and $\Phi = \{\varphi, \mu\}$, as follows:

$$\begin{aligned} A'(U)(\Phi, Z) &= \lambda\, m(\varphi, z) - a(\varphi, z) + \mu\, m(u, z) + 2\bar{\pi}\mathrm{Re}\, m(\varphi, u), \\ J'(U)(\Phi) &= \mu\, m(u, u) + 2\lambda\mathrm{Re}\, m(\varphi, u). \end{aligned}$$

Hence, the continuous dual problem takes the form

$$\lambda\, m(\varphi, z) - a(\varphi, z) + \mu\{m(u, z) - m(u, u)\} + 2\{\bar{\pi} - \lambda\}\mathrm{Re}\, m(\varphi, u) \;=\; 0,$$

for all $\Phi = \{\varphi, \mu\}$. We now show that the dual eigenpair $U^* = \{u^*, \lambda^*\}$ solves this equation, i.e., $Z = \{z, \pi\} = U^*$ is one solution. This is clear since for $m(u, z) = m(u, u) = 1$ and $\bar{\pi} = \lambda$, the equation reduces to

$$a(\varphi, z) = \bar{\pi}\, m(\varphi, z),$$

which is the defining equation for U^*. Whether this is the only solution is not relevant for the present discussion. Analogously, the discrete adjoint problem (7.8) is solved by the discrete dual eigenpair $U_h^* = \{u_h^*, \lambda_h^*\}$ determined by

$$a(\varphi_h, u_h^*) = \bar{\lambda}_h^*\, m(\varphi_h, u_h^*) \quad \forall \varphi_h \in V_h, \quad m(u_h, u_h^*) = 1. \tag{7.9}$$

With these preparations, we can formulate the following proposition:

Proposition 7.1. *With the primal and dual eigenvalue residuals*

$$\rho(u_h, \lambda_h)(\cdot) := a(u_h, \cdot) - \lambda_h\, m(u_h, \cdot),$$
$$\rho^*(u_h^*, \lambda_h^*)(\cdot) := a(\cdot, u_h^*) - \bar{\lambda}_h^*\, m(\cdot, u_h^*),$$

we have the error representation

$$\lambda - \lambda_h \;=\; \tfrac{1}{2}\rho(u_h, \lambda_h)(u^* - \psi_h) + \tfrac{1}{2}\rho^*(u_h^*, \lambda_h^*)(u - \varphi_h) \;+\; \mathcal{R}_h, \tag{7.10}$$

for arbitrary ψ_h, $\varphi_h \in V_h$, with the cubic remainder term

$$\mathcal{R}_h \;=\; \tfrac{1}{2}(\lambda - \lambda_h)\, m(u - u_h, u^* - u_h^*).$$

Proof. The assertion is an immediate consequence of Proposition 6.2 applied to the present situation. We have

$$\begin{aligned} J(U) - J(U_h) \;=\; & \tfrac{1}{2}\big\{J'(U_h)(U - \Phi_h) - A'(U_h)(U - \Phi_h, Z_h)\big\} \\ & + \tfrac{1}{2}\big\{ - A(U_h)(Z - \Psi_h)\big\} + \mathcal{R}_h, \end{aligned}$$

for arbitrary $\Phi_h = \{\varphi_h, \mu\}$, $\Psi_h = \{\psi_h, \chi\} \in \mathcal{V}_h$, with the cubic remainder term $\mathcal{R}_h$ which we evaluate below. Hence, using the above preparations,

$$\begin{aligned} \lambda - \lambda_h \;=\; & \tfrac{1}{2}\big\{(\lambda - \mu)\, m(u_h, u_h) + 2\lambda_h \operatorname{Re} m(u - \varphi_h, u_h) \\ & + a(u - \varphi_h, z_h) - \lambda_h\, m(u - \varphi_h, z_h) - (\lambda - \mu)\, m(u_h, z_h) \\ & - 2\bar{\pi}_h \operatorname{Re} m(u - \varphi_h, u_h) - (\lambda - \mu)\{m(u_h, u_h) - 1\}\big\} \\ & - \tfrac{1}{2}\big\{\lambda_h\, m(u_h, z - \psi_h) - a(u_h, z - \psi_h) + (\bar{\pi} - \bar{\chi})\,\{m(u_h, u_h) - 1\}\big\} \\ & + \mathcal{R}_h. \end{aligned}$$

We are free to choose $\mu := \lambda$, $\chi := \pi$, and observing that $m(u, u) = m(u_h, u_h) = 1$, and $\lambda_h = \bar{\pi}_h = \bar{\lambda}_h^*$, we obtain

$$\begin{aligned} \lambda - \lambda_h \;=\; & \tfrac{1}{2}\big\{a(u - \varphi_h, z_h) - \lambda_h\, m(u - \varphi_h, z_h)\big\} \\ & + \tfrac{1}{2}\big\{a(u_h, z - \psi_h) - \bar{\lambda}_h^*\, m(u_h, z - \psi_h)\big\} \;+\; \mathcal{R}_h. \end{aligned}$$

It remains to evaluate the remainder term. Setting $E := \{u - u_h, \lambda - \lambda_h\}$ and $E^* := \{u^* - u_h^*, \lambda^* - \lambda_h^*\}$, the general remainder term from Proposition 6.2 is

$$\begin{aligned} \mathcal{R}_h \;=\; \tfrac{1}{2}\int_0^1 \big\{ & J'''(U_h + sE)(E, E, E) - A'''(U_h + sE)(E, E, E, Z_h + sE^*) \\ & - 3A''(U_h + sE)(E, E, E^*)\big\} s(s - 1)\, ds. \end{aligned}$$

In the present case, by a simple calculation, we have

$$\begin{aligned} J'''(U_h+sE)(E,E,E) &= 6(\lambda-\lambda_h)m(u-u_h,u-u_h), \\ A'''(U_h+sE)(E,E,E,Z_h+sE^*) &= 0, \\ -3A''(U_h+sE)(E,E,E^*) &= -6(\lambda-\lambda_h)\,m(u-u_h,u^*-u_h^*) \\ &\quad - 6(\bar\lambda^*-\bar\lambda_h^*)\,m(u-u_h,u-u_h). \end{aligned}$$

Consequently, noting that $\lambda-\lambda_h = \bar\lambda^*-\bar\lambda_h^*$,

$$\begin{aligned} \mathcal{R}_h &= -3\int_0^1 (\lambda-\lambda_h)\,m(u-u_h,u^*-u_h^*)s(s-1)\,ds \\ &= \tfrac{1}{2}(\lambda-\lambda_h)\,m(u-u_h,u^*-u_h^*), \end{aligned}$$

which completes the proof. □

Remark 7.2. We add the following remarks concerning Proposition 7.1:

- The error representation (7.10) holds true without any assumption on the multiplicity of the eigenvalue λ or its defect. However, such a restriction will become necessary below in dealing with the error of the eigenfunctions.

- In the error representation (7.10) only terms involving the computed primal and dual eigenpairs occur and no additional outer dual problem needs to be solved. We will see a similar situation in the context of optimization problems discussed in Chapter 8 where the underlying mechanism will become clear.

- In the nonsymmetric case the simultaneous consideration of primal and dual eigenvalue problems is essential within an optimal multigrid iteration anyway (see Heuveline and Bertsch [76]). Then, the computation of u^* for the error estimator does therefore not introduce extra work.

- In Proposition 7.1, we have assumed that the governing operator $\mathcal{A}$ remains unchanged under discretization, i.e. all coefficient are frozen. Below, in considering the stability eigenvalue problem, we will additionally allow for approximation of the operator $\mathcal{A}(\hat u)$ depending on coefficients $\hat u$ that will also be subject to approximation.

Practical evaluation of the error representation

Next, we want to determine the explicit form of the residuals in Proposition 7.1. To this end, we need to be more specific about the particular structure of the eigenvalue problem considered. Here, we restrict ourselves to a simple model situation which, nevertheless, is prototypical for the problems we are interested in. On a polygonal or polyhedral domain $\Omega \subset \mathbb{R}^d$ consider the eigenvalue problem of a second-order elliptic differential operator $\mathcal{A}$ such as, for example,

$$\mathcal{A}v := -\Delta v + b\cdot\nabla v = \lambda v \quad \text{in } \Omega, \quad \mathrm{v}_{|\partial\Omega} = 0,$$

with a smooth (or even constant) transport coefficient b. In this case, we have $\mathcal{M} := \mathrm{id}$. Further, let this eigenvalue problem be approximated by the Galerkin method using piecewise linear or d-linear finite elements on meshes $\mathbb{T}_h = \{K\}$, as described in Chapter 3. Within this setting, we can proceed analogously as before, obtaining

$$\begin{aligned}\rho(u_h,\lambda_h)(\cdot) &= a(u_h,\cdot) - \lambda_h\, m(u_h,\cdot)\\ &= \sum_{K\in\mathbb{T}_h} \left\{(\mathcal{A}u_h - \lambda_h\mathcal{M}u_h,\cdot)_K - (\partial_n^{\mathcal{A}}u_h,\cdot)_{\partial K}\right\}\\ &= \sum_{K\in\mathbb{T}_h} \left\{(\mathcal{A}u_h - \lambda_h\mathcal{M}u_h,\cdot)_K + \tfrac{1}{2}([\partial_n^{\mathcal{A}}u_h],\cdot)_{\partial K}\right\},\end{aligned}$$

$$\begin{aligned}\rho^*(u_h^*,\lambda_h^*)(\cdot) &= a(\cdot,z_h) - \lambda_h^*\, m(\cdot,z_h)\\ &= \sum_{K\in\mathbb{T}_h} \left\{(\cdot,\mathcal{A}^*z_h - \lambda_h^*\mathcal{M}z_h)_K - (\cdot,\partial_n^{\mathcal{A}^*}z_h)_{\partial K}\right\}\\ &= \sum_{K\in\mathbb{T}_h} \left\{(\cdot,\mathcal{A}^*z_h - \lambda_h^*\mathcal{M}z_h)_K + \tfrac{1}{2}(\cdot,[\partial_n^{\mathcal{A}^*}z_h])_{\partial K}\right\}.\end{aligned}$$

Hence, using again the notation of 'equation' and 'jump residuals' R_h, R_h^*, r_h, and r_h^*, respectively, analogously as introduced in Section 3.1, the residual admits the estimate

$$|\rho(u_h,\lambda_h)(u^*-\psi_h) + \rho^*(u_h^*,\lambda_h^*)(u-\varphi_h)| \le \sum_{K\in\mathbb{T}_h} \left\{\rho_K\omega_K^* + \rho_K^*\omega_K\right\}, \tag{7.11}$$

with the cell residuals ρ_K, ρ_K^* and and weights ω_K, ω_K^* defined by

$$\begin{aligned}\rho_K &:= \left(\|R_h\|_K^2 + h_K^{-1/2}\|r_h\|_{\partial K}^2\right)^{1/2},\\ \rho_K^* &:= \left(\|R_h^*\|_K^2 + h_K^{-1/2}\|r_h^*\|_{\partial K}^2\right)^{1/2},\\ \omega_K &:= \left(\|u-\varphi_h\|_K^2 + \tfrac{1}{2}h_K^{1/2}\|u-\varphi_h\|_{\partial K}^2\right)^{1/2},\\ \omega_K^* &:= \left(\|u^*-\psi_h\|_K^2 + h_K^{1/2}\|u^*-\psi_h\|_{\partial K}^2\right)^{1/2}.\end{aligned}$$

As a consequence of the above discussion, we obtain the following result:

Proposition 7.3. *Within the above setting, assuming that*

$$|m(u-u_h,u^*-u_h^*)| \le 1, \tag{7.12}$$

we have the 'weighted' a posteriori error estimate

$$|\lambda-\lambda_h| \le \eta_\lambda^\omega := \sum_{K\in\mathbb{T}_h} \left\{\rho_K\omega_K^* + \rho_K^*\omega_K\right\}, \tag{7.13}$$

and the 'energy-norm-error-type' estimate

$$|\lambda-\lambda_h| \le \eta_\lambda^{(1)} := c_\lambda \sum_{K\in\mathbb{T}_h} h_K^2 \{\rho_K^2 + \rho_K^{*2}\}, \tag{7.14}$$

with a constant c_λ *growing linearly with* $|\lambda|$.

Proof. Using the estimate (7.11) in the error representation (7.10) gives us

$$|\lambda-\lambda_h| \le \tfrac{1}{2} \sum_{K\in\mathbb{T}_h} \{\rho_K \omega_K^* + \rho_K^* \omega_K\} + |\mathcal{R}_h|.$$

Since, in virtue of assumption (7.12),

$$|\mathcal{R}_h| = \tfrac{1}{2}|(\lambda-\lambda_h)\, m(u-u_h, u^*-u_h^*)| \le \tfrac{1}{2}\,|\lambda-\lambda_h|,$$

the asserted estimate (7.13) follows. To prove (7.14), we choose $\psi_h := \tilde{I}_h u^*$ and $\varphi_h := \tilde{I}_h u$ in ω_K^* and ω_K, with the modified nodal interpolation operator $\tilde{I}_h$ introduced in Section 3.2. Writing

$$u^*-\tilde{I}_h u^* = (u^*-u_h^*) - \tilde{I}_h(u^*-u_h^*), \quad u-\tilde{I}_h u = (u-u_h) - \tilde{I}_h(u-u_h),$$

in the weights ω_K^* and ω_K, we obtain by the interpolation estimate (3.10) that

$$\sum_{K\in\mathbb{T}_h} h_K^{-2}\{\omega_K^{*2} + \omega_K^2\} \le \tilde{c}_I^2 \{\|\nabla(u^*-u_h^*)\|^2 + \|\nabla(u-u_h)\|^2\}.$$

This gives us

$$|\lambda-\lambda_h| \le \tfrac{1}{2}\tilde{c}_I \Big(\sum_{K\in\mathbb{T}_h} h_K^2 \{\rho_K^2 + \rho_K^{*2}\}\Big)^{1/2} \big(\|\nabla(u^*-u_h^*)\|^2 + \|\nabla(u-u_h)\|^2\big)^{1/2}.$$

Now, the proof would be completed by showing that the energy-norm errors of the eigenfunctions are proportionally bounded by the eigenvalue error. This is actually the case but requires lengthy calculations employing duality arguments for the eigenfunction errors. These details are omitted and we refer instead to Heuveline and Rannacher [77]. □

Remark 7.4. The analogue of the a posteriori error estimator $\eta_\lambda^{(1)}$ for *symmetric* eigenvalue problems has been given by Nystedt [108]. There, the symmetry of the problem is extensively used. Furthermore, H^2-regularity of the eigenfunctions is required which excludes domains with reentrant corners, the most interesting case for an adaptive approach. Notice that our derivation of $\eta_\lambda^{(1)}$ does not need these assumptions.

Remark 7.5. An alternative version of the eigenvalue-error estimator $\eta_\lambda^{(1)}$ has been given by Larson [101], also for the symmetric case and assuming H^2-regularity of the eigenfunctions, namely,

$$|\lambda - \lambda_h| \leq \eta_\lambda^{(2)} := c_\lambda \Big(\sum_{K \in \mathbb{T}_h} h_K^4 \{ \rho_K^2 + \rho_K^{*2} \} \Big)^{1/2}. \tag{7.15}$$

Both error estimators, $\eta_\lambda^{(1)}$ and $\eta_\lambda^{(2)}$, are asymptotically equivalent for regular situations. However, the required H^2 regularity for the eigenfunctions renders the estimator $\eta_\lambda^{(2)}$ useless for typical situations in which mesh adaptivity is needed.

Remark 7.6. Neglecting the presence of the *dual* eigenvalue problem, we may try to control the error in approximating the eigenvalue using the *primal* residual part alone, i.e. using the 'reduced' error estimator

$$\eta_\lambda^{\mathrm{red}} := \sum_{K \in \mathbb{T}_h} h_K^2 \rho_K^2.$$

Below, we will compare the performance of the above eigenvalue-error estimators η_λ^ω, $\eta_\lambda^{(1)}$, $\eta_\lambda^{(2)}$, and $\eta_\lambda^{\mathrm{red}}$ for a simple model situation.

Numerical examples

We consider the convection-diffusion model eigenvalue problem

$$-\Delta v + b \cdot \nabla v \;=\; \lambda v \quad \text{in } \Omega, \quad \mathrm{v}_{|\partial\Omega} = 0,$$

on the slit-domain $\Omega = (-1,1)\times(-1,3) \setminus \{x \in \mathbb{R}^2, x_1 = 0, -1 < x_2 \leq 0\}$, with the transport vector $b = (0, b_y)^T$; for a sketch of this configuration, see Figure 7.1. In the computations on this test problem the mesh refinement is organized according to the 'fixed-rate' strategy with refinement rate $X = 0.2$, see Section 4.2.

Test 1: Symmetric case

At first, we consider the symmetric eigenvalue problem, i.e. $b = 0$. The eigenfunction corresponding to the smallest eigenvalue and the computational mesh generated on the basis of the weighted error estimator η_λ^ω are shown in Figure 7.1. Figure 7.2 shows the mesh efficiencies achieved on the basis of the different error estimators η_λ^ω, $\eta_\lambda^{(1)}$ and $\eta_\lambda^{(2)}$ introduced above compared with that of uniform mesh refinement. (In this case $\eta_\lambda^{\mathrm{red}}$ and $\eta_\lambda^{(1)}$ are equivalent.) We see that in the symmetric case all estimators show almost equally good performance compared to that of uniform refinement. This is due to the dominance of the error caused by the slit singularity which is well captured by all residual-based error estimators. In fact, it is known that in the symmetric case, the eigenvalue error is proportional to the square of the energy-norm error.

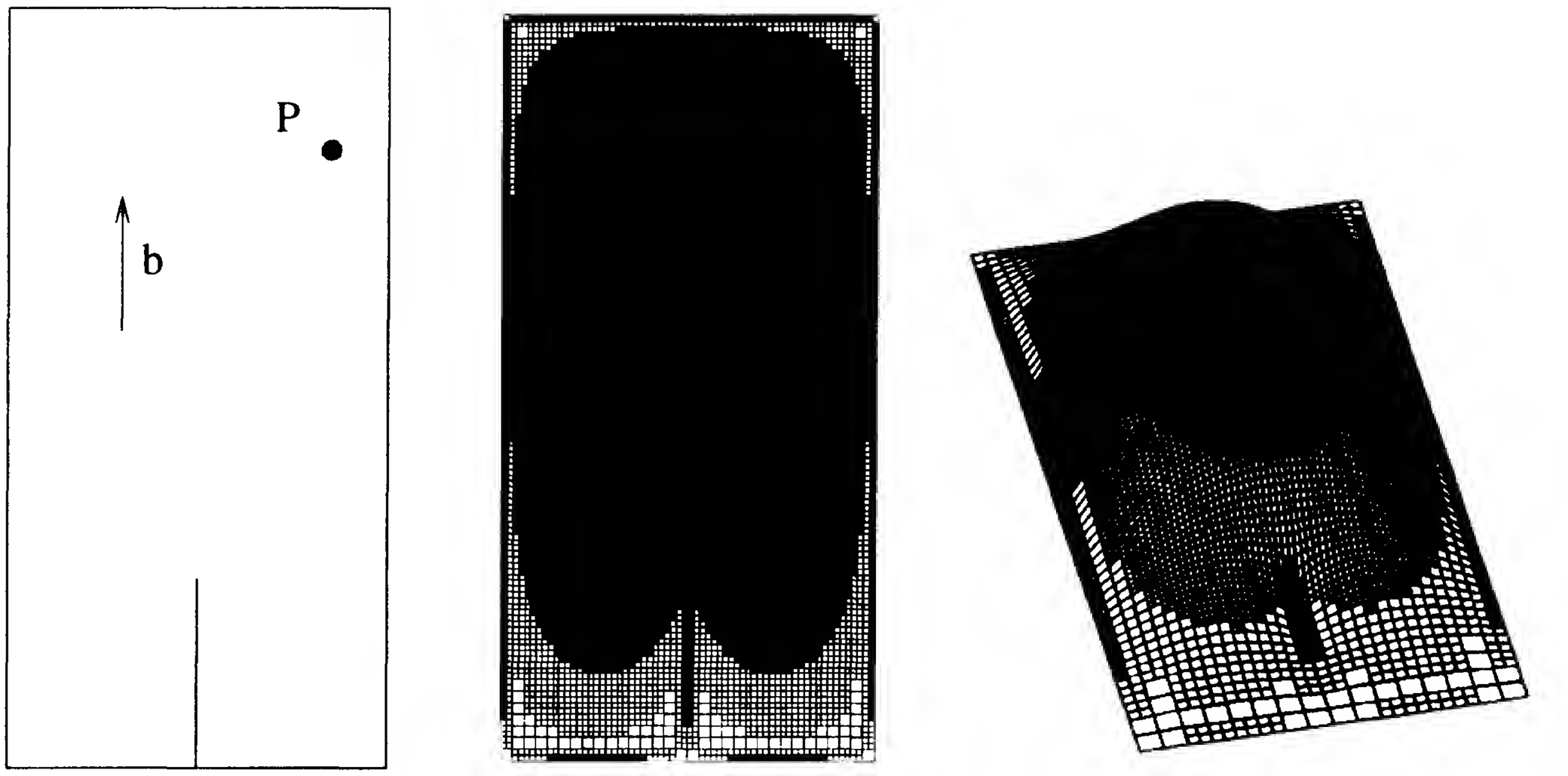


Figure 7.1: *Configuration for* $b \equiv 0$ *(left), adapted mesh with* $12,000$ *cells (middle), eigenfunction (right); from Heuveline and Rannacher [77].*

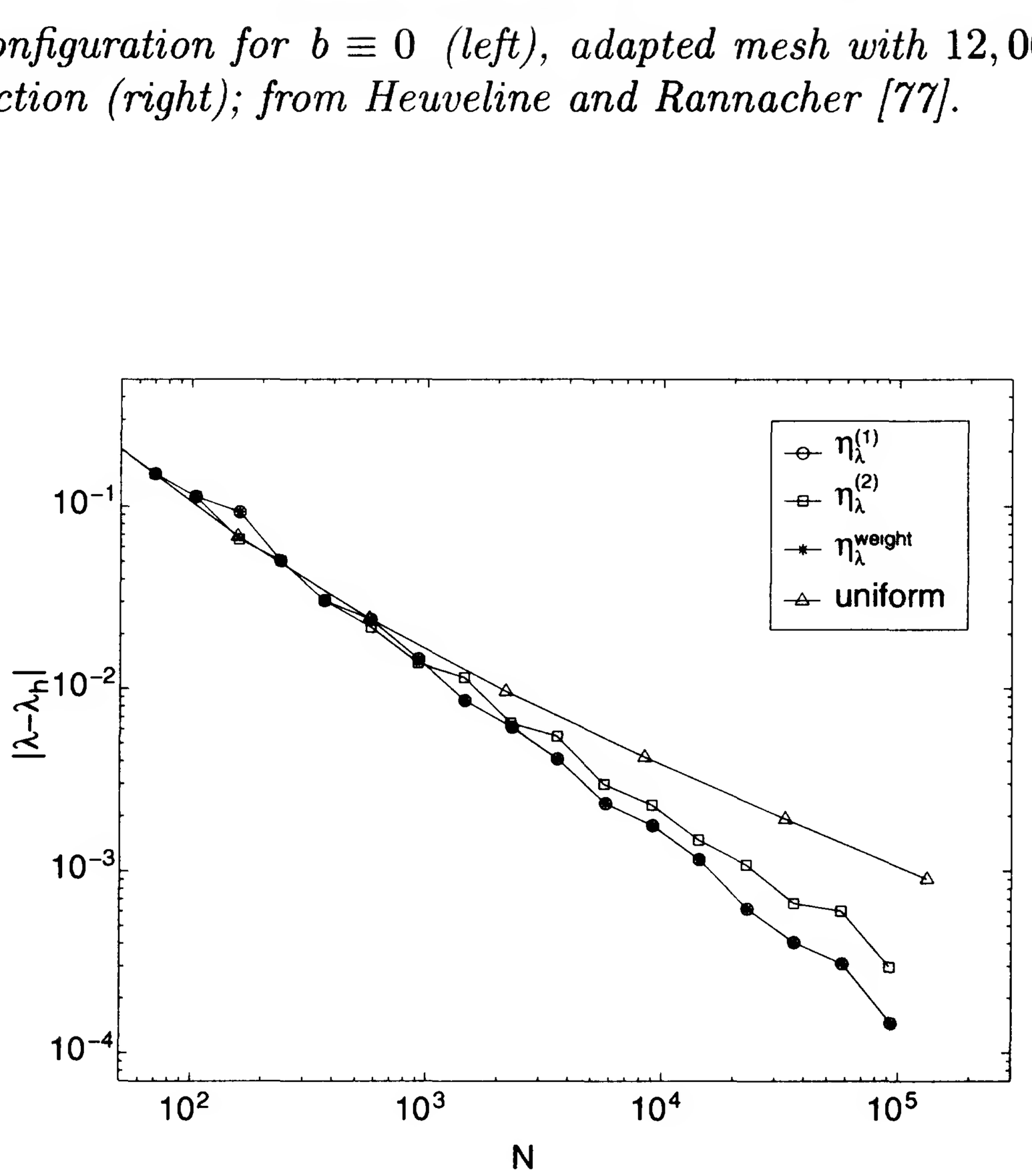


Figure 7.2: *Mesh efficiency achieved using the different error estimators,* $\eta_\lambda^{(1)}$ *(symbol* $\circ$*),* $\eta_\lambda^{(2)}$ *(symbol* $\square$*), and* η_λ^{ω} *(symbol* $*$*), compared against uniform refinement (symbol* $\triangle$*). The curves for* $\eta_\lambda^{(1)}$ *and* η_λ^{ω} *lie above each other; from Heuveline and Rannacher [77].*

Test 2: Nonsymmetric case

Next, we consider the nonsymmetric version of the test eigenvalue problem with vertical transport, $b_y = 3$. In this case, due to the Dirichlet boundary conditions, the primal eigenfunction has a steeper gradient at the top boundary, while the dual eigenfunction has one at the bottom boundary; see Figure 7.3. The latter boundary layer strongly interferes with the slit singularity. Figure 7.4 shows adapted meshes obtained using the 'energy-type' error estimator $\eta_\lambda^{(1)}$, the 'reduced' error estimator η_λ^{red}, and the 'weighted' error estimator η_λ^{ω}. The superiority of the latter one is clearly seen in Figure 7.5.

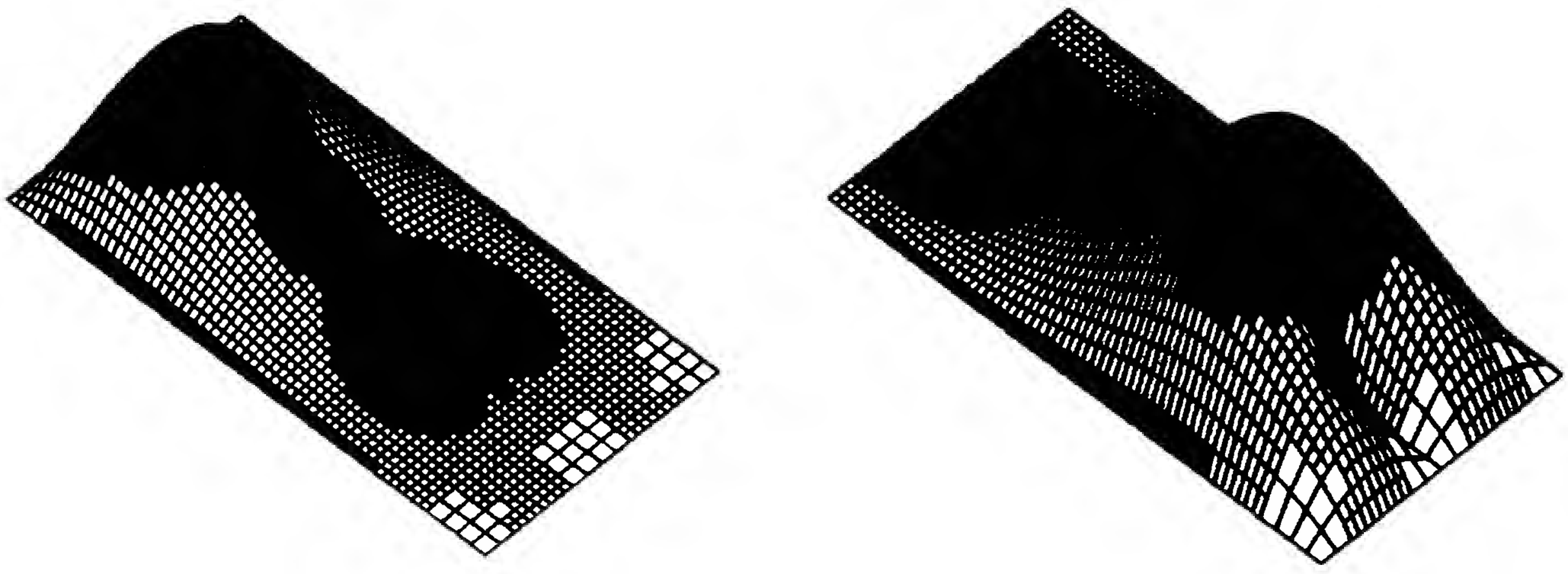


Figure 7.3: *Primal (left) and dual eigenfunction (right); from Heuveline and Rannacher [77].*

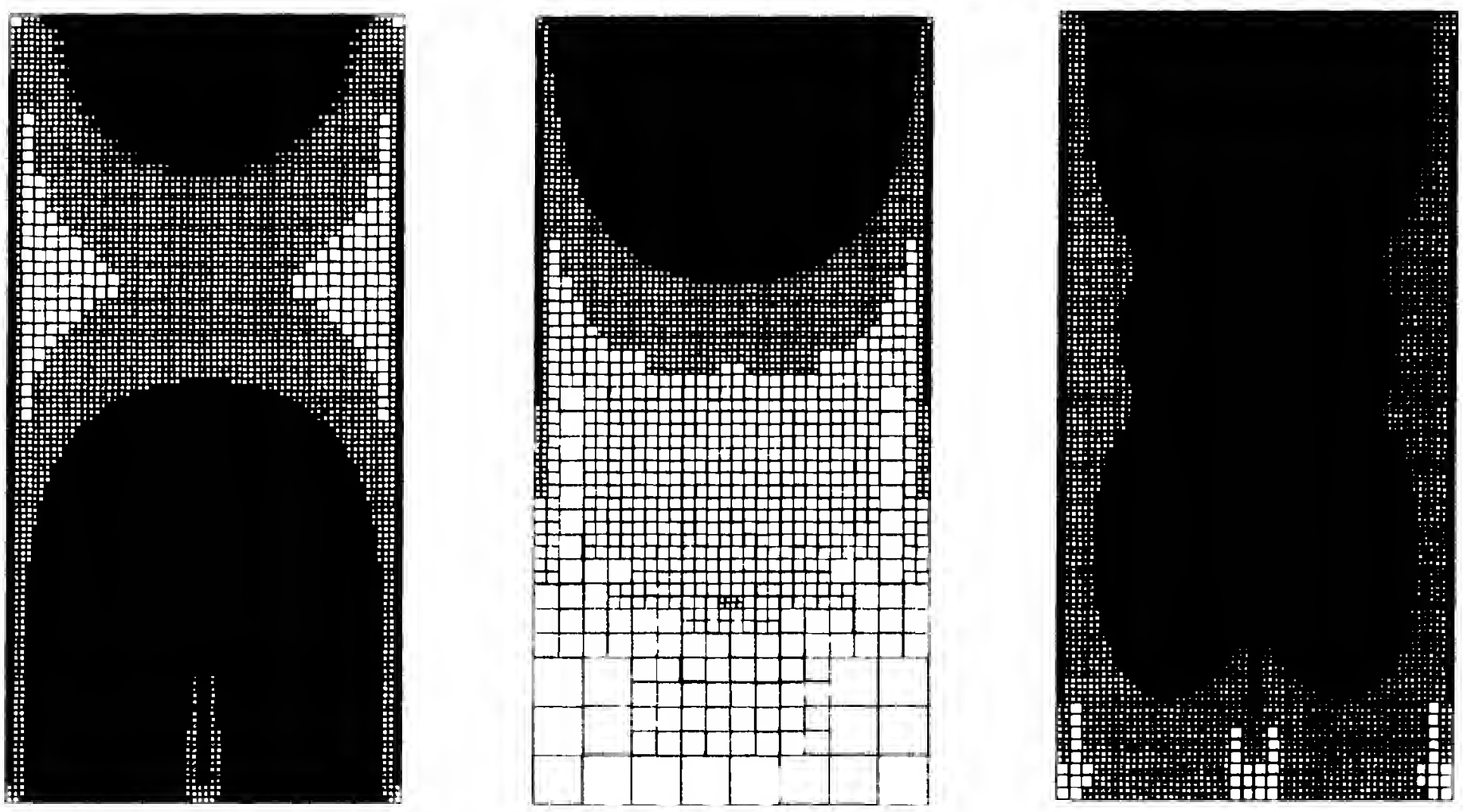


Figure 7.4: *Adapted meshes with* $10,000$ *cells on the basis of the error estimators* $\eta_\lambda^{(1)}$ *(left),* η_λ^{red} *(middle),* η_λ^{ω} *(right); from Heuveline and Rannacher [77].*

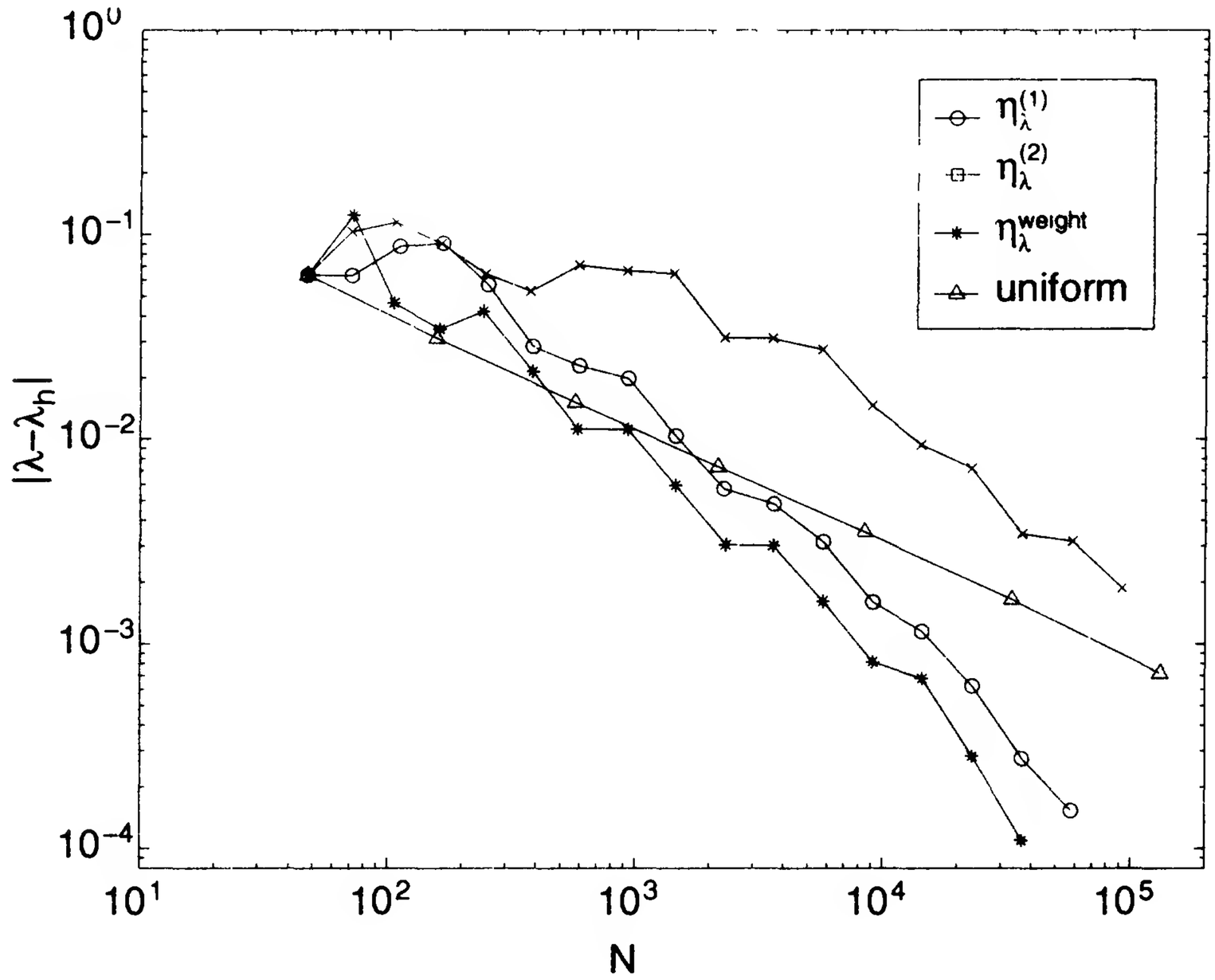


Figure 7.5: *Mesh efficiencies achieved on the basis of the different error estimators:* $\eta_\lambda^{(1)}$ *(symbol* $\circ$*),* η_λ^{red} *(symbol* $\times$*), and* η_λ^{ω} *(symbol* $*$*), compared against uniform refinement (symbol* $\triangle$*); from Heuveline and Rannacher [77].*

7.2 Error control for functionals of eigenfunctions

We now turn to the question of how to estimate the error with respect to output functionals of the eigenfunctions. This question seems non-trivial since at an eigenvalue the corresponding adjoint operator is naturally singular, so that it is not clear what the proper definition of the dual problem has to be in this case.

Let $j(\cdot) : V \to \mathbb{C}$ be an output functional, for simplicity assumed to be linear, with respect to which the error in the eigenfunctions is to be controlled. For ease of presentation, we also assume that the eigenvalue λ considered is simple with a normalized eigenfunction u. In order to apply our abstract theory, we define the functional

$$J(\Phi) := j(\varphi), \quad \Phi = \{\varphi, \mu\} \in \mathcal{V}.$$

Then, the associated abstract dual problem for $Z = \{z, \pi\} \in \mathcal{V}$,

$$A'(U)(\Phi, Z) = J'(U)(\Phi) \qquad \forall \Phi \in \mathcal{V},$$

takes the explicit form

$$\lambda\, m(\varphi, z) - a(\varphi, z) + \mu\, m(u, z) - 2\,\bar{\pi}\ \mathrm{Re}\ m(\varphi, u) \;=\; j(\varphi), \tag{7.16}$$

for all $\Phi = \{\varphi, \mu\} \in \mathcal{V}$. This is equivalent to

$$\lambda\, m(\varphi, z) - a(\varphi, z) \;=\; j(\varphi) + 2\,\bar{\pi}\ \mathrm{Re}\ m(\varphi, u) \quad \forall \varphi \in V,$$

and $m(u, z) = 0$. Since λ is assumed to be a simple eigenvalue, by virtue of the Fredholm alternative, this equation can be solved if and only if its right-hand side vanishes on the eigenvector u, that is

$$j(u) - 2\,\bar{\pi}\ \mathrm{Re}\, m(u, u) = 0 \quad \Leftrightarrow \quad \bar{\pi} = \tfrac{1}{2}\, j(u).$$

Consequently, for the given eigenpair $\{u, \lambda\}$ the so reduced dual problem

$$a(\varphi, z) - \lambda\, m(\varphi, z) \;=\; j(\varphi) - j(u)\ \mathrm{Re}\ m(\varphi, u) \quad \forall \varphi \in V, \tag{7.17}$$

has a solution $z \in V$, which is uniquely determined in view of the property $m(u, z) = 0$. By an analogous argument, the reduced discrete dual problem is seen to be

$$a(\varphi_h, z_h) - \lambda_h\, m(\varphi_h, z_h) \;=\; j(\varphi_h) - j(u_h)\ \mathrm{Re}\ m(\varphi_h, u_h) \quad \forall \varphi_h \in V_h, \tag{7.18}$$

where $\{u_h, \lambda_h\}$ is the eigenpair of the approximate eigenvalue problem, and $\bar{\pi}_h := \frac{1}{2}\, j(u_h)$. This problem also has a solution $z_h \in V_h$, uniquely determined by $m(u_h, z_h) = 0$. We see that the well-posedness of these dual problems is guaranteed by filtering out the eigenspaces $\mathrm{span}\{u\}$ and $\mathrm{span}\{u_h\}$, respectively. With all this, we can state the following proposition.

Proposition 7.7. *Let $\{u_h, \lambda_h\}$ be a computed eigenpair approximating $\{u, \lambda\}$. Then, for the given functional $j(\cdot) : V \to \mathbb{C}$ and the associated solution $z \in V$ of the dual problem (7.17), we have the error representation*

$$\begin{aligned} j(u - u_h) \;=\; & \rho(u_h, \lambda_h)(z - \psi_h) + (\lambda - \lambda_h)\, m(u - u_h, z) \\ & + \tfrac{1}{2} j(u)\, m(u - u_h, u - u_h), \end{aligned} \tag{7.19}$$

for arbitrary $\psi_h \in V_h$.

Proof. First, we recall the definitions of the primal (eigenvalue) residual

$$\rho(u_h, \lambda_h)(\cdot) = a(u_h, \cdot) - \lambda_h m(u_h, \cdot),$$

and the dual residual associated to the dual problem (7.17),

$$\rho^*(u_h, z_h)(\cdot) := \; a(\cdot, z_h) - \lambda_h\, m(\cdot, z_h) + j(\cdot) - j(u_h)\, \mathrm{Re}\, m(\cdot, u_h).$$

This implies with the results of Chapter 6 that

$$\begin{aligned} j(u - u_h) \;=\; & \tfrac{1}{2} \left\{ a(u_h, z - \psi_h) - \lambda_h m(u_h, z - \psi_h) \right\} \\ & + \tfrac{1}{2} \left\{ a(u - \varphi_h, z_h) - \lambda_h m(u - \varphi_h, z_h) \right. \\ & \left. - j(u_h)\, \mathrm{Re}\ m(u - \varphi_h, u_h) + j(u - \varphi_h) \right\} + \mathcal{R}_h\,, \end{aligned}$$

and consequently, taking $\varphi_h = u_h$,

$$\begin{aligned} j(u-u_h) &= a(u_h, z-\psi_h) - \lambda_h m(u_h, z-\psi_h) + a(u-u_h, z_h) \\ &\quad - \lambda_h m(u-u_h, z_h) - j(u_h)\,\mathrm{Re}\; m(u-u_h, u_h) + 2\mathcal{R}_h\,. \end{aligned}$$

To identify the remainder $\mathcal{R}_h$, we note that

$$\begin{aligned} J'''(U_h+sE; E, E, E) &= 0, \\ A'''(U_h+sE; Z_h+sE^*)(E, E, E) &= 0, \\ -3A''(U_h+sE; E, E, E^*) &= -6(\lambda-\lambda_h)m(u-u_h, z-z_h) \\ &\quad - 6(\bar\pi-\bar\pi_h)m(u-u_h, u-u_h), \end{aligned}$$

which yields

$$\mathcal{R}_h = \tfrac{1}{2}(\lambda-\lambda_h)m(u-u_h, z-z_h) + \tfrac{1}{2}(\bar\pi-\bar\pi_h)m(u-u_h, u-u_h).$$

We recall that $\bar\pi = \frac{1}{2}\, j(u)$ and $\bar\pi_h = \frac{1}{2}\, j(u_h)$ and obtain

$$\mathcal{R}_h = \tfrac{1}{2}(\lambda-\lambda_h)m(u-u_h, z-z_h) + \tfrac{1}{4}\, j(u-u_h)m(u-u_h, u-u_h).$$

From this, we infer as an intermediate result:

$$\begin{aligned} j(u-u_h) &= a(u_h, z-\psi_h) - \lambda_h m(u_h, z-\psi_h) \\ &\quad + a(u-u_h, z_h) - \lambda_h m(u-u_h, z_h) \\ &\quad - j(u_h)\,\mathrm{Re}\; m(u-u_h, u_h) + \tfrac{1}{2} j(u-u_h)m(u-u_h, u-u_h) \\ &\quad + (\lambda-\lambda_h)m(u-u_h, z-z_h). \end{aligned}$$

Next, by definition and since $m(u_h, z_h) = 0$, we have

$$a(u-u_h, z_h) - \lambda_h m(u-u_h, z_h) = (\lambda-\lambda_h)m(u, z_h) = (\lambda-\lambda_h)m(u-u_h, z_h).$$

Further, noting that $m(u, u) = m(u_h, u_h) = 1$, there holds

$$\begin{aligned} m(u-u_h, u-u_h) &= m(u, u) + m(u_h, u_h) - 2\,\mathrm{Re}\; m(u, u_h) \\ &= m(u, u) - m(u_h, u_h) - 2\,\mathrm{Re}\; m(u-u_h, u_h) \\ &= -2\,\mathrm{Re}\; m(u-u_h, u_h). \end{aligned}$$

Then, combining the last three relations gives us

$$\begin{aligned} j(u-u_h) &= a(u_h, z-\psi_h) - \lambda_h m(u_h, z-\psi_h) + (\lambda-\lambda_h)m(u-u_h, z_h) \\ &\quad + \tfrac{1}{2} j(u)\, m(u-u_h, u-u_h) + (\lambda-\lambda_h)m(u-u_h, z-z_h), \end{aligned}$$

which completes the proof. □

Remark 7.8. The proposition requires λ to be simple. In the case of geometric multiplicity $\rho > 1$, we have to simultaneously consider a whole basis $\{u^{(i)}, i = 1, ..., \rho\}$ of the eigenspace $\mathrm{kern}(\mathcal{A} - \lambda I)$ in setting up the dual problem. The case of higher algebraic multiplicity can also be handled but is much more involved.

Numerical example

In order to illustrate the foregoing result, we consider the model problem from above. The goal is to evaluate the derivative value $j(u) := \partial_1 u(a)$ of the 'first' eigenfunction at $a = (0.5, 2.5)^T$.

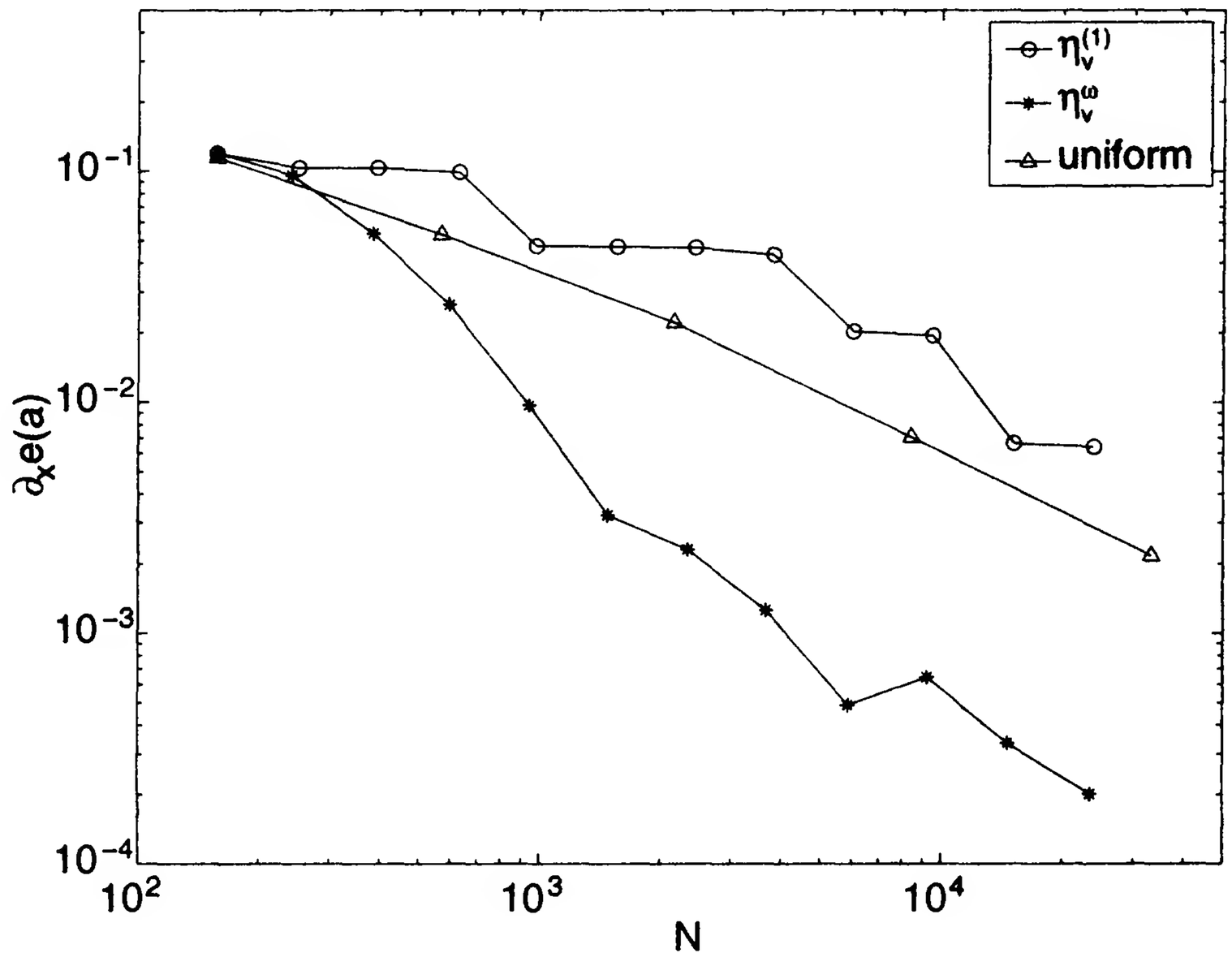


Figure 7.6: *Mesh efficiency of η_u^{ω} (symbol $*$) compared to $\eta_u^{(1)}$ (symbol O) and uniform refinement (symbol Δ); from Heuveline and Rannacher [77].*

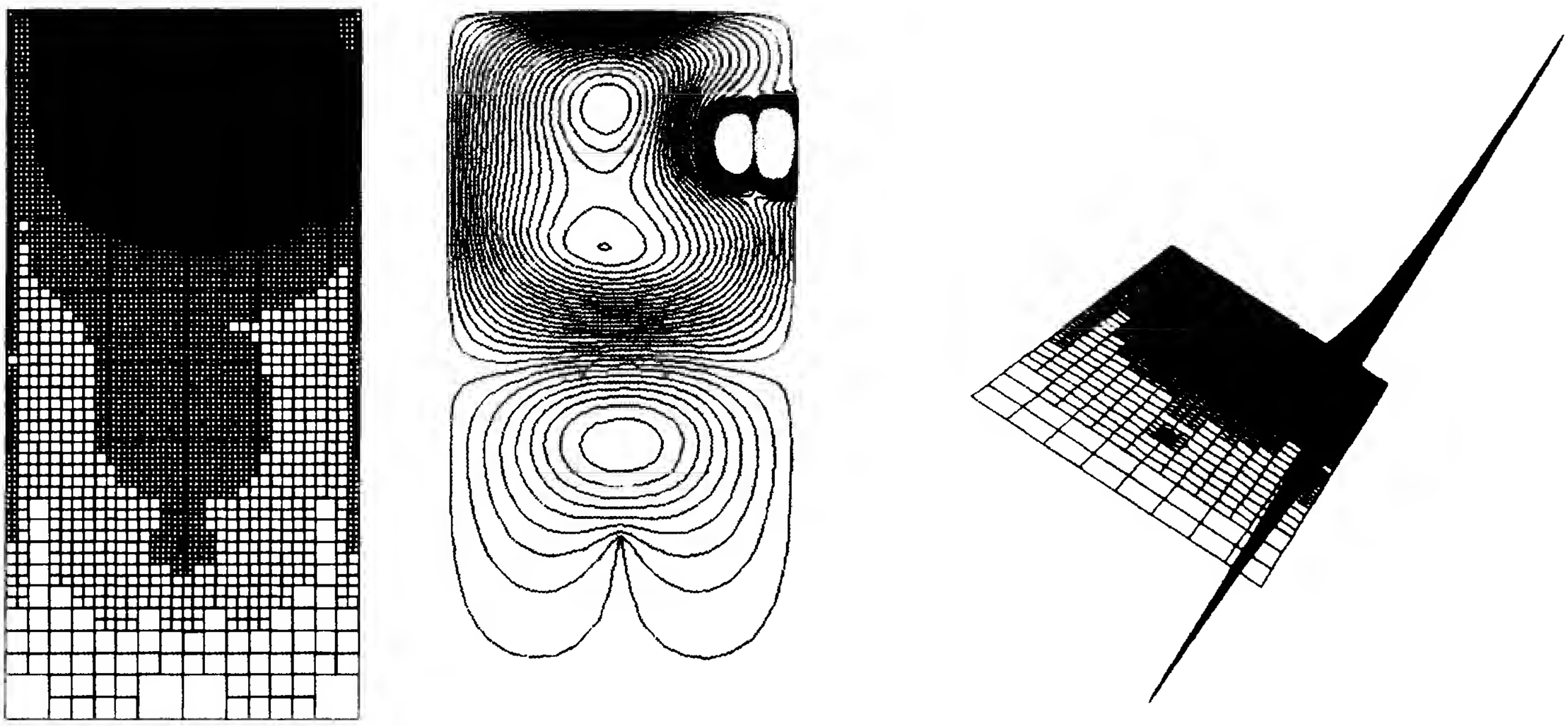


Figure 7.7: *Adapted mesh with about* $12,000$ *cells for the approximation of* $\partial_1 u(a)$ *obtained by* η_u^{ω} *(left), and the corresponding dual solution (middle and right); from Heuveline and Rannacher [77].*

7.3 The stability eigenvalue problem

Eigenvalue problems play an important role in the analysis of the *stability* of solutions of nonlinear differential equations. Classical areas of applications are structural mechanics and fluid mechanics, in the latter context referred to as *hydrodynamic stability*. We will consider this kind of problem again within an abstract setting. Let $a(\cdot)(\cdot)$ be a given semilinear form determining the base solution $\hat{u} \in V$ by

$$a(\hat{u})(\psi) = 0 \quad \forall \psi \in V, \tag{7.20}$$

and its Galerkin approximation $\hat{u}_h \in V_h \subset V$ by

$$a(\hat{u}_h)(\psi_h) = 0 \quad \forall \psi_h \in V_h. \tag{7.21}$$

Note that $\hat{u}$ is a *stationary* solution of the system under consideration. We want to determine whether this base solution is *dynamically stable*, i.e., whether any non-stationary solution trajectory $\{\tilde{u}(t) \in V, t \geq 0\}$, starting from a small perturbation $\tilde{u}(0) = \hat{u} + w^0$, and satisfying the evolution equation associated to (7.20),

$$m(\partial_t \tilde{u}, \psi) + a(\tilde{u})(\psi) = 0 \quad \forall \psi \in V, \tag{7.22}$$

stays bounded or even decays back to $\hat{u}$, as $t \to \infty$. Here, the decay is expressed in the semi-norm $m(\cdot,\cdot)^{1/2}$, which in most practical cases is actually the usual L^2 norm. However, in order to prepare for the application of this approach also to the incompressible Navier-Stokes equations in Chapter 11, we allow $m(\cdot,\cdot)$ to be slightly more general.

In the following, we outline the basics of the so-called *linearized stability* theory. The perturbation $w(t) := \tilde{u}(t) - \hat{u}$ satisfies the nonlinear *perturbation equation*

$$\begin{aligned} 0 &= m(\partial_t w, \psi) + a(\hat{u}+w)(\psi) - a(\hat{u})(\psi) \\ &= m(\partial_t w, \psi) + a'(\hat{u})(w, \psi) + \int_0^1 a''(\hat{u}+sw)(w, w, \psi) s \, ds, \end{aligned}$$

for $t \geq 0$. Taking $\psi = w$, we obtain

$$\frac{1}{2}\frac{d}{dt} m(w, w) + a'(\hat{u})(w, w) = -\int_0^1 a''(\hat{u}+sw)(w, w, w) s \, ds.$$

Assuming now that the perturbation $w(t)$ is initially small such the the cubic term on the right can be neglected, the initial decay or growth of $w(t)$ is determined by the spectral properties of the sesquilinear form $a'(\hat{u})(\cdot,\cdot)$ obtained by linearizing $a(\hat{u})(\cdot)$ at the base solution. Particularly, if the nonsymmetric *stability eigenvalue problem*

$$a'(\hat{u})(u, \psi) = \lambda m(u, \psi) \quad \forall \psi \in V, \tag{7.23}$$

has an eigenvalue λ^{crit} with negative real part, then any perturbation having initially a nontrivial component in the direction of an eigenfunction of λ^{crit} will grow initially and may eventually blow up.

We want to solve this stability eigenvalue problem numerically by a Galerkin approximation which reads as follow:

$$a'(\hat{u}_h)(u_h, \psi_h) = \lambda_h m(u_h, \psi_h) \quad \forall \psi_h \in V_h. \tag{7.24}$$

Our goal is to estimate the error in the critical eigenvalue, $\lambda^{crit} - \lambda_h^{crit}$, i.e. the 'first' eigenvalue with possibly negative real part, in terms of the residuals corresponding to primal and dual equations. Note that this includes both the error due to approximating the base solution $\hat{u}$, as well as of the eigenvalue. To this end, we embed this situation into the general framework of variational equations laid out above. We introduce the spaces

$$\mathcal{V} := V \times V \times \mathbb{C}, \quad \mathcal{V}_h := V_h \times V_h \times \mathbb{C},$$

consisting of elements $U := \{\hat{u}, u, \lambda\}$, $\Psi = \{\hat{\psi}, \psi, \nu\}$ and $U_h := \{\hat{u}_h, u_h, \lambda_h\}$, $\Psi_h = \{\hat{\psi}_h, \psi_h, \nu\}$, respectively. Using the semilinear form

$$A(U)(\Psi) := a(\hat{u})(\hat{\psi}) + \lambda\, m(u, \psi) - a'(\hat{u})(u, \psi) + \overline{\nu}\big\{m(u, u) - 1\big\},$$

the continuous and discrete problems can be written in compact form as follows:

$$A(U)(\Psi) = 0 \qquad \forall \Psi \in \mathcal{V}, \tag{7.25}$$

$$A(U_h)(\Psi_h) = 0 \qquad \forall \Psi_h \in \mathcal{V}_h, \tag{7.26}$$

For controling the eigenvalue error, we work again with the functional

$$J(\Phi) := \mu\, m(\varphi, \varphi), \qquad \Phi = \{\hat{\varphi}, \varphi, \mu\},$$

such that $J(U) = \lambda m(u, u) = \lambda$. Then, the corresponding continuous and discrete dual solutions $Z = \{\hat{z}, z, \pi\} \in \mathcal{V}$ and $Z_h = \{\hat{z}_h, z_h, \pi_h\} \in \mathcal{V}_h$ are determined by the problems

$$A'(U)(\Phi, Z) = J'(U)(\Phi) \qquad \forall \Phi \in \mathcal{V}, \tag{7.27}$$

$$A'(U_h)(\Phi_h, Z_h) = J'(U_h)(\Phi_h) \qquad \forall \Phi_h \in \mathcal{V}_h. \tag{7.28}$$

A detailed computation shows that these equations are solved by $Z = U^* := \{\hat{z}, u^*, \lambda\}$ and $Z_h = U_h^* := \{\hat{z}_h, u_h^*, \lambda_h\}$, respectively, where u^* and u_h^* are the (normalized) dual eigenfunctions as before, and the dual base solutions $\hat{z}$ and $\hat{z}_h$ are determined by the dual problems

$$a'(\hat{u})(\varphi, \hat{z}) = -a''(\hat{u})(\varphi, u, u^*) \qquad \forall \varphi \in V, \tag{7.29}$$

$$a'(\hat{u}_h)(\varphi_h, \hat{z}_h) = -a''(\hat{u}_h)(\varphi_h, u_h, u_h^*) \qquad \forall \varphi \in V. \tag{7.30}$$

The corresponding residuals of the approximate base solutions are

$$\begin{aligned}\hat{\rho}(\hat{u}_h)(\cdot) &:= -a(\hat{u}_h)(\cdot),\\ \hat{\rho}^*(\hat{u}_h,\hat{z}_h)(\cdot) &:= a''(\hat{u}_h)(\cdot,u_h,u_h^*) - a'(\hat{u}_h)(\cdot,\hat{z}_h),\end{aligned}$$

and those of the eigenpair approximations,

$$\begin{aligned}\rho(\hat{u}_h,u_h,\lambda_h)(\cdot) &:= a'(\hat{u}_h)(u_h,\cdot) - \lambda_h\, m(u_h,\cdot),\\ \rho^*(\hat{u}_h,u_h^*,\lambda_h^*)(\cdot) &:= a'(\hat{u}_h)(\cdot,u_h^*) - \bar{\lambda}_h^*\, m(\cdot,u_h^*).\end{aligned}$$

We collect the foregoing findings in the following proposition:

Proposition 7.9. *With the above residuals, we have the error representation*

$$\begin{aligned}\lambda-\lambda_h = {}& \tfrac{1}{2}\hat{\rho}(\hat{u}_h)(\hat{z}-\hat{\psi}_h) + \tfrac{1}{2}\hat{\rho}^*(\hat{u}_h,\hat{z}_h)(\hat{u}-\hat{\varphi}_h)\\ &+ \tfrac{1}{2}\rho(\hat{u}_h,u_h,\lambda_h)(u^*-\psi_h) + \tfrac{1}{2}\rho^*(\hat{u}_h,u_h^*,\lambda_h^*)(u-\varphi_h) \;+\; \mathcal{R}_h,\end{aligned} \tag{7.31}$$

for arbitrary $\hat{\psi}_h, \psi_h, \hat{\varphi}_h, \varphi_h \in V_h$. *The remainder* $\mathcal{R}_h$ *is cubic in the errors* $\hat{e}^u := \hat{u}-\hat{u}_h$, $\hat{e}^z := \hat{z}-\hat{z}_h$, *and* $e^\lambda := \lambda-\lambda_h$, $e^u := u-u_h$, $e^{u*} := u^*-u_h^*$:

$$\begin{aligned}\mathcal{R}_h \;=\; \tfrac{1}{2}e^\lambda m(e^u,e^{u*}) + \tfrac{1}{2}\int_0^1 \Big\{ & a''''(\hat{u}_h+s\hat{e}^u)(\hat{e}^u,\hat{e}^u,\hat{e}^u,u_h+se^u,u_h^*+se^{u*})\\ & - a'''(\hat{u}_h+s\hat{e}^u)(\hat{e}^u,\hat{e}^u,\hat{e}^u,\hat{z}_h+s\hat{e}^z)\\ & + 3a'''(\hat{u}_h+s\hat{e}^u)(\hat{e}^u,\hat{e}^u,e^u,u_h^*+se^{u*})\\ & + 3a'''(\hat{u}_h+s\hat{e}^u)(\hat{e}^u,\hat{e}^u,u_h+se^u,e^{u*})\\ & + 6a''(\hat{u}_h+s\hat{e}^u)(\hat{e}^u,e^u,e^{u*})\\ & - 3a''(\hat{u}_h+s\hat{e}^u)(\hat{e}^u,\hat{e}^u,\hat{e}^{u*})\Big\} s(s-1)\,ds.\end{aligned}$$

Application to a model case

We want to apply this abstract result to the concrete situation of the nonsymmetric model problem from Example 6.1 (vector Burgers equation). Here, the stability eigenvalue problem has the form

$$-\Delta u + \hat{u}\cdot\nabla u + u\cdot\nabla\hat{u} \;=\; \lambda u, \quad \text{in } \Omega, \quad u_{|\partial\Omega} = 0. \tag{7.32}$$

Its finite element approximation computes the discrete base solution $\hat{u}_h$ by solving

$$(\nabla\hat{u}_h,\nabla\psi_h) + (\hat{u}_h\cdot\nabla\hat{u}_h,\psi_h) = (f,\psi_h) \quad \forall\psi_h\in V_h, \tag{7.33}$$

and then determines the corresponding eigenvalue λ_h from the eigenvalue problem

$$(\nabla u_h,\nabla\psi_h) + (\hat{u}_h\cdot\nabla u_h + u_h\cdot\nabla\hat{u}_h,\psi_h) \;=\; \lambda_h\,(u_h,\psi_h) \quad \forall\psi_h\in V_h. \tag{7.34}$$

The corresponding discrete dual problem determining $\hat{z}_h \in V_h$ reads

$$\begin{aligned}(\nabla\varphi_h, \nabla\hat{z}_h) + (\varphi_h, \nabla\hat{u}_h\hat{z}_h - \nabla\cdot(\hat{u}_h \otimes \hat{z}_h)) \\ = (\varphi_h, \nabla u_h u_h^* - \nabla\cdot(u_h \otimes u_h^*)) \quad \forall\varphi_h \in V_h,\end{aligned} \tag{7.35}$$

and that determining the discrete dual eigenpair $\{u_h^*, \lambda_h^*\}$,

$$(\nabla\varphi_h, \nabla u_h^*) + (\varphi_h, \nabla\hat{u}_h u_h^* - \nabla\cdot(\hat{u}_h \otimes u_h^*)) = \bar{\lambda}_h^*(\varphi_h, u_h^*) \quad \forall\varphi_h \in V_h. \tag{7.36}$$

Then, the associated cell residuals are

$$\begin{aligned}\hat{R}_{h|K} &:= f + \Delta\hat{u}_h - \hat{u}_h\cdot\nabla\hat{u}_h, \\ \hat{R}_{h|K}^* &:= \Delta\hat{z}_h + \nabla u_h u_h^* - \nabla\cdot(u_h \otimes u_h^*) - \nabla\hat{u}_h\hat{z}_h + \nabla\cdot(\hat{u}_h \otimes \hat{z}_h), \\ R_{h|K} &:= -\Delta u_h + \hat{u}_h\cdot\nabla u_h + u_h\cdot\nabla\hat{u}_h - \lambda_h u_h, \\ R_{h|K}^* &:= -\Delta u_h^* - \nabla\cdot(\hat{u}_h \otimes u_h^*) + \nabla\hat{u}_h u_h^* - \bar{\lambda}_h^* u_h^*,\end{aligned}$$

while the associated edge residuals $\hat{r}_{h|\Gamma}$, $\hat{r}_{h|\Gamma}^*$ and $r_{h|\Gamma}$, $r_{h|\Gamma}^*$ have the same form as above in the case of the simple Poisson equation. For this situation Proposition 7.9 yields the following result:

Proposition 7.10. *Using the notation from above, and assuming again that*

$$|m(u-u_h, u^*-u_h^*)| \le 1,$$

we have the a posterior error estimate

$$|\lambda-\lambda_h| \le \eta_\lambda^\omega := \sum_{K\in\mathbb{T}_h} \left\{\hat{\rho}_K\hat{\omega}_K^* + \hat{\rho}_K^*\hat{\omega}_K + \rho_K\omega_K^* + \rho_K^*\omega_K\right\} + \mathcal{R}_h. \tag{7.37}$$

The cell residuals and weights are defined by

$$\begin{aligned}\hat{\rho}_K &:= \left(\|\hat{R}_h\|_K^2 + h_K^{-1/2}\|\hat{r}_h\|_{\partial K}^2\right)^{1/2}, \\ \hat{\omega}_K^* &:= \left(\|\hat{z}-\hat{\psi}_h\|_K^2 + h_K^{1/2}\|\hat{z}-\hat{\psi}_h\|_{\partial K}^2\right)^{1/2}, \\ \hat{\rho}_K^* &:= \left(\|\hat{R}_h^*\|_K^2 + h_K^{-1/2}\|\hat{r}_h^*\|_{\partial K}^2\right)^{1/2}, \\ \hat{\omega}_K &:= \left(\|\hat{u}-\hat{\varphi}_h\|_K^2 + \tfrac{1}{2}h_K^{1/2}\|\hat{u}-\hat{\varphi}_h\|_{\partial K}^2\right)^{1/2}, \\ \rho_K &:= \left(\|R_h\|_K^2 + h_K^{-1/2}\|r_h\|_{\partial K}^2\right)^{1/2}, \\ \omega_K^* &:= \left(\|u^*-\psi_h\|_K^2 + h_K^{1/2}\|u^*-\psi_h\|_{\partial K}^2\right)^{1/2}, \\ \rho_K^* &:= \left(\|R_h^*\|_K^2 + h_K^{-1/2}\|r_h^*\|_{\partial K}^2\right)^{1/2}, \\ \omega_K &:= \left(\|u-\varphi_h\|_K^2 + \tfrac{1}{2}h_K^{1/2}\|u-\varphi_h\|_{\partial K}^2\right)^{1/2}.\end{aligned}$$

for arbitrary $\hat{\psi}_h, \psi_h, \hat{\varphi}_h, \varphi_h \in V_h$, and the remainder $\hat{\mathcal{R}}_h$ is cubic in the errors $\hat{e}^u := \hat{u}-\hat{u}_h$ and $\hat{e}^{u*} := \hat{u}^*-\hat{u}_h^*$:

$$\mathcal{R}_h = -\left(\nabla(\hat{e}^u \otimes e^u - \tfrac{1}{2}\hat{e}^u\hat{\otimes}e^u), e^{u*}\right).$$

An error estimator as derived in Proposition 7.10 will be applied below in Chapter 11 to the approximation of the stability eigenvalue problem of the Navier-Stokes equations. There, we will demonstrate the interplay of the different components in the error estimator η_λ^ω.

7.4 Exercises

Exercise 7.1. For the Galerkin approximation of a symmetric eigenvalue problem

$$a(u,\psi) = \lambda\,(u,\psi) \quad \forall \psi \in V,$$

with a symmetric bilinear form $a(\cdot,\cdot)$ on a (real) Hilbert space V, the general a posteriori error representation reduces to the form

$$\lambda - \lambda_h \;=\; \rho(u_h,\lambda_h)(u-\psi_h) + \tfrac{1}{2}(\lambda-\lambda_h)\|u-u_h\|^2, \quad \psi \in V_h.$$

Derive this identity by direct algebraic manipulation.

Exercise 7.2. Consider the model convection-diffusion eigenvalue problem

$$-\Delta v + b\cdot\nabla v \;=\; \lambda v \quad \text{in } \Omega, \quad v_{|\partial\Omega} = 0.$$

Under the assumption that the eigenfunctions have H^2-regularity, and that

$$|(u-u_h, u^*-u_h^*)| \le 1,$$

prove the a posteriori error bound

$$|\lambda-\lambda_h| \;\le\; \eta_\lambda^{(2)} := c_\lambda \Big(\sum_{K\in\mathbb{T}_h} h_K^4 \{\rho_K^2 + \rho_K^{*2}\} \Big)^{1/2},$$

with a constant $c_\lambda = \mathcal{O}(|\lambda|)$.

Exercise 7.3. Consider the d-dimensional Burgers equation

$$-\nu\Delta u + u\cdot\nabla u = f \;\text{ in } \Omega, \quad u_{|\partial\Omega} = 0,$$

for a vector function $u : \Omega \to \mathbb{R}^d$. Suppose that $\hat{u} \in V := H_0^1(\Omega)^d$ is a solution. Show that if all eigenvalues of the *symmetric* stability eigenvalue problem

$$-\nu\Delta w + \tfrac{1}{2}\{\nabla\hat{u} + \nabla\hat{u}^T - \nabla\cdot\hat{u}I\}w = \lambda w$$

are positive, then $\hat{u}$ is *dynamically stable.* This criterion for stability is much stronger than that of *linearized stability* since it allows perturbations of any size and also applies to nonstationary solutions. Therefore, it cannot be expected to be satisfied in many practically interesting situations.

Exercise 7.4 *(Practical exercise).* Consider the nonlinear boundary value problem of Exercise 5.3,

$$-\Delta u - u^3 = f, \text{ in } \Omega, \quad u_{|\partial\Omega} = 0,$$

where Ω is the domain defined in Exercise 3.3. The discretization is by the usual bilinear finite elements. For $f \equiv \alpha = 1, \ldots, 75$, compute the solution on increasingly refined meshes and investigate its dynamic stability using the criterion of *linearized stability*. Monitor the behavior of the eigenvalue error estimator and, in particular, its components measuring the approximation of the base solution and that of the eigenvalue. Explain the observed results.

Bibliography

[1] M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods. *Numer. Math.*, 65:23–50, 1993.

[2] M. Ainsworth and J. T. Oden. A posteriori error estimation in finite element analysis. *Comput. Methods Appl. Mech. Eng.*, 142:1–88, 1997.

[3] M. Ainsworth and J. T. Oden. A posteriori error estimators for the Stokes and Oseen equations. *SIAM J. Numer. Anal.*, 34:228–245, 1997.

[4] J. P. Aubin. Behaviour of the error of the approximate solutions of boundary value problems for linear elliptic operators by Galerkin's and finite difference methods. *Ann. Scuola Morm. Sup. Pisa*, 21:599–637, 1967.

[5] I. Babuška and A. D. Miller. The post-processing approach in the finite element method, I: calculations of displacements, stresses and other higher derivatives of the displacements. *Int. J. Numer. Meth. Eng.*, 20:1085–1109, 1984.

[6] I. Babuška and A. D. Miller. The post-processing approach in the finite element method, II: the calculation of stress intensity factors. *Int. J. Numer. Meth. Eng.*, 20:1111–1129, 1984.

[7] I. Babuška and A. D. Miller. The post-processing approach in the finite element method, III: a posteriori error estimation and adaptive mesh selection. *Int. J. Numer. Meth. Eng.*, 20:2311–2324, 1984.

[8] I. Babuška and A. D. Miller. A feedback finite element method with a posteriori error estimation. *Comput. Methods Appl. Mech. Eng.*, 61:1–40, 1987.

[9] I. Babuška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.*, 15:736–754, 1978.

[10] I. Babuška and W. C. Rheinboldt. A posteriori error estimates for the finite element method. *Int. J. Numer. Meth. Eng.*, 12:1597–1615, 1978.

[11] I. Babuška and C. Schwab. A posteriori error estimation for hierarchic models of elliptic boundary value problems on thin domains. *SIAM J. Numer. Anal.*, 33:221–246, 1996.

[12] I. Babuška and T. Strouboulis. *The Finite Element Method and its Reliability.* Clarendon Press, Oxford, 2001.

[13] E. Backes. Gewichtete a posteriori Fehleranalyse bei der adaptiven Finite-Elemente-Methode: Ein Vergleich zwischen Residuen- und Bank-Weiser-Schätzer. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 1997.

[14] W. Bangerth. Finite Element Approximation of the Acoustic Wave Equation: Error Control and Mesh Adaptation. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 1998.

[15] W. Bangerth. Adaptive Finite Element Methods for the Identification of Distributed Parameters in Partial Differential Equations. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 2002.

[16] W. Bangerth and R. Rannacher. Finite element approximation of the acoustic wave equation: Error control and mesh adaptation. *East–West J. Numer. Math.*, 7:263–282, 1999.

[17] R. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations. *Math. Comp.*, 44:283–301, 1985.

[18] R. Becker. An Adaptive Finite Element Method for the Incompressible Navier–Stokes Equations on Time-Dependent Domains. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 1995.

[19] R. Becker. An adaptive finite element method for the Stokes equations including control of the iteration error. *ENUMATH'97* (H. G. Bock *et al.*, eds), pp. 609–620, World Scientific, Singapore, 1998.

[20] R. Becker. An optimal-control approach to a posteriori error estimation for finite element discretizations of the Navier-Stokes equations. *East-West J. Numer. Math.*, 9:257–274, 2000.

[21] R. Becker. Mesh adaptation for stationary flow control. *J. Math. Fluid Mech.*, 3:317–341, 2001.

[22] R. Becker. Adaptive Finite Elements for Optimal Control Problems. Habilitation thesis, University of Heidelberg, 2001.

[23] R. Becker and M. Braack. Multigrid techniques for finite elements on locally refined meshes. *Numer. Linear Algebra Appl.*, 7:363–379, 2000.

[24] R. Becker and M. Braack. Solution of a stationary benchmark problem for natural convection with large temperature difference. *Int. J. Therm. Sci.*, 41:428–439, 2002.

[25] R. Becker, M. Braack, and R. Rannacher. Numerical simulation of laminar flames at low Mach number by adaptive finite elements. *Combust. Theory Modelling*, 3:503–534, 1999.

[26] R. Becker, M. Braack, R. Rannacher, and C. Waguet. Fast and reliable solution of the Navier–Stokes equations including chemistry. *Comput. Visual. Sci.*, 2:107–122, 1999.

[27] R. Becker, C. Johnson, and R. Rannacher. Adaptive error control for multigrid finite element methods. *Computing*, 55:271–288, 1995.

[28] R. Becker, H. Kapp, and R. Rannacher. Adaptive finite element methods for optimal control of partial differential equations: Basic concepts. *SIAM J. Control Optim.*, 39, 113–132, 2000.

[29] R. Becker and R. Rannacher. Weighted a posteriori error control in FE methods. Lecture at ENUMATH-95, Paris, Sept. 18-22, 1995, Preprint 96-01, SFB 359, University of Heidelberg, Proc. *ENUMATH'97* (H. G. Bock *et al.*, eds), pp. 621-637, World Scientific, Singapore, 1998.

[30] R. Becker and R. Rannacher. A feed-back approach to error control in finite element methods: Basic analysis and examples. *East-West J. Numer. Math.*, 4:237–264, 1996.

[31] R. Becker and R. Rannacher. An optimal control approach to error estimation and mesh adaptation in finite element methods. *Acta Numerica 2000* (A. Iserles, ed.), pp. 1-101, Cambridge University Press, 2001.

[32] R. Becker and B. Vexler. Adaptive finite element methods for parameter identification problems. Preprint 2002-20 (SFB 359), Universität Heidelberg, July 2002, *Numer. Math.*, submitted, 2002

[33] C. Bernardi, O. Bonnon, C. Langouët, and B. Métivet. Residual error indicators for linear problems: Extension to the Navier–Stokes equations. In Proc. *9th Int. Conf. Finite Elements in Fluids*, 1995.

[34] H. Blum, Q. Lin, and R. Rannacher. Asymptotic error expansion and Richardson extrapolation for linear finite elements. *Numer. Math.*, 49:11–37, 1986.

[35] H. Blum and F.-T. Suttmeier. An adaptive finite element discretization for a simplified Signorini problem. *Calcolo*, 37:65–77, 1999.

[36] H. Blum and F.-T. Suttmeier. Weighted error estimates for finite element solutions of variational inequalities. *Computing*, 65:119–134, 2000.

[37] K. Böttcher. Adaptive Schrittweitenkontrolle beim unstetigen Galerkin-Verfahren für gewöhnliche Differentialgleichungen. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 1996.

[38] K. Böttcher and R. Rannacher. Adaptive error control in solving ordinary differential equations by the discontinuous Galerkin method. Technical Report Preprint 96-53, SFB 359, Universität Heidelberg, 1996.

[39] M. Braack. An Adaptive Finite Element Method for Reactive Flow Problems. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 1998.

[40] M. Braack, R. Becker, and R. Rannacher. The dual-weighted residual method applied to three-dimensional flow problems. Computers & Fluids, Proc. 3. Int. Conf. on Appl. Math. for Industrial Flow Problems (AMIF), Lisbon, April 17-20, 2002, to appear.

[41] M. Braack and A. Ern. A posteriori control of modeling errors and discretization errors. Preprint 2002-13 (SFB 359), Universität Heidelberg, June 2002, SIAM J. Multiscale Modeling and Simulation, submitted, 2002.

[42] M. Braack and R. Rannacher. Adaptive finite element methods for low-Mach-number flows with chemical reactions. In H. Deconinck, editor, *30th Computational Fluid Dynamics, Lecture Series*, Vol. 1999-03, von Karman Institute for Fluid Dynamics, 1999.

[43] S. Brenner and R. L. Scott. *The Mathematical Theory of Finite Element Methods.* Springer, Berlin Heidelberg New York, 1994.

[44] G. F. Carey and J. T. Oden. *Finite Elements, Computational Aspects*, Vol. III. Prentice-Hall, 1984.

[45] C. Carstensen and R. Verfürth. Edge residuals dominate a posteriori error estimators for low-order finite element methods. *SIAM J. Numer. Anal.*, 36:1571–1587, 1999.

[46] P. G. Ciarlet. *Finite Element Methods for Elliptic Problems.* North-Holland, Amsterdam, 1978.

[47] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc.*, 49:1–23, 1943.

[48] W. Dörfler and M. Rumpf. An adaptive strategy for elliptic problems including a posteriori controlled boundary approximation. *Math. Comp.*, 224:1361–1382, 1998.

[49] T. Dunne. Adaptive dual-gemischte Finite-Elemente-Verfahren. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 2001.

[50] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. Introduction to adaptive methods for differential equations. *Acta Numerica 1995* (A. Iserles, ed.), pp. 105–158, Cambridge University Press, 1995.

[51] K. Eriksson and C. Johnson. An adaptive finite element method for linear elliptic problems. *Math. Comp.*, 50:361–383, 1988.

[52] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems, I: A linear model problem. *SIAM J. Numer. Anal.*, 28:43–77, 1991.

[53] K. Eriksson and C. Johnson. Adaptive streamline diffusion finite element methods for stationary convection-diffusion problems. *Math. Comp.*, 60:167–188, 1993.

[54] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems, II: Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. *SIAM J. Numer. Anal.*, 32:706–740, 1995.

[55] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems, IV: Nonlinear problems. *SIAM J. Numer. Anal.*, 32:1729–1749, 1995.

[56] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems, V: Long-time integration. *SIAM J. Numer. Anal.*, 32:1750–1763, 1995.

[57] K. Eriksson, C. Johnson, and S. Larsson. Adaptive finite element methods for parabolic problems, VI: Analytic semigroups. *SIAM J. Numer. Anal.*, 35:1315–1325, 1998.

[58] D. Estep. A posteriori error bounds and global error control for approximation of ordinary differential equations. *SIAM J. Numer. Anal.*, 32:1–48, 1995.

[59] D. Estep and D. French. Global error control for the continuous Galerkin finite element method for ordinary differential equations. *Modél. Math. Anal. Numér.*, 28:815–852, 1994.

[60] J. Frehse and R. Rannacher. Eine L^1-Fehlerabschätzung für diskrete Grundlösungen in der Methode der finiten Elemente. Tagungsband Finite Elemente, *Bonn. Math. Schr.*, 89:92–114, 1976.

[61] C. Führer. Error Control in Finite Element Methods for Hyperbolic Problems. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 1996.

[62] C. Führer and G. Kanschat. A posteriori error control in radiative transfer. *Computing*, 58:317–334, 1997.

[63] C. Führer and R. Rannacher. An adaptive streamline-diffusion finite element method for hyperbolic conservation laws. *East–West J. Numer. Math.*, 5:145–162, 1997.

[64] M. B. Giles. On adjoint equations for error analysis and optimal grid adaptation. In Frontiers of Computational Fluid Dynamics 1998 (D.A. Caughey and M.M. Hafez, eds), pp. 155-170. , World Scientific, 1998.

[65] M. B. Giles and N. A. Pierce. Adjoint error correction for integral outputs. Technical Report NA-01/18, Oxford University Computing Laboratory, 2001.

[66] M. B. Giles, M. G. Larsson, J. M. Levenstam, and E. Süli. Adaptive error control for finite element approximations of the lift and drag coefficients in viscous flow. NA-97/06, Oxford University Computing Laboratory, 1997.

[67] P. Hansbo. Three lectures on error estimation and adaptivity. *Adaptive Finite Elements in Linear and Nonlinear Solid and Structural Mechanics* (E. Stein, ed.), Vol. 416 of CISM Courses and Lectures, Springer, 2002, to appear.

[68] P. Hansbo and C. Johnson. Adaptive streamline diffusion finite element methods for compressible flow using conservative variables. *Comput. Methods Appl. Mech. Eng.*, 87:267–280, 1991.

[69] R. Hartmann. A posteriori Fehlerschätzung und adaptive Schrittweiten- und Ortsgittersteuerung bei Galerkin-Verfahren für die Wärmeleitungsgleichung. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 1998.

[70] R. Hartmann. Adaptive FE-methods for conservation equations. In Proc. *8th International Conference on Hyperbolic Problems. Theory, Numerics, Applications (HYP2000)* (H. Freistühler and G. Warnecke, eds.), pp. 495–503, Int. Series of Numer. Math. 141, Birkhäuser, Basel, 2001.

[71] R. Hartmann. Adaptive Finite Element Methods for the Compressible Euler Equations. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 2002.

[72] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. Preprint 2001-20 (SFB 359), Universität Heidelberg, *SIAM J. Sci. Comput.*, to appear.

[73] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for the compressible Euler equations. Preprint 2001-41 (SFB 359), Universität Heidelberg, *J. Comput. Physics*, to appear.

[74] F.-K. Hebeker and R. Rannacher. An adaptive finite element method for unsteady convection-dominated flows with stiff source terms. *SIAM J. Sci. Comput.*, 21:799–818, 1999.

[75] J. Heywood, R. Rannacher, and S. Turek. Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations. *Int. J. Comput. Fluid Mech.*, 22:325–352, 1996.

[76] V. Heuveline and C. Bertsch. *On multigrid methods for the eigenvalue computation of non-selfadjoint elliptic operators.* East-West J. Numer. Math. 8, 275–297 (2000).

[77] V. Heuveline and R. Rannacher. A posteriori error control for finite element approximations of elliptic eigenvalue problems. Preprint 2001-08 (SFB 359), University of Heidelberg, *J. Comput. Math. Appl.*, 15:107–138, 2001.

[78] V. Heuveline and R. Rannacher. Adaptive finite element discretization of eigenvalue problems in hydrodynamic stability theory. Preprint, SFB 359, Universität Heidelberg, March 2001.

[79] P. Houston, R. Rannacher, and E. Süli. A posteriori error analysis for stabilized finite element approximation of transport problems. *Comput. Methods Appl. Mech. Eng.*, 190:1483–1508, 2000.

[80] T. J. R. Hughes and A. N. Brooks. Streamline upwind/Petrov Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equation. *Comput. Methods Appl. Mech. Eng.*, 32:199–259, 1982.

[81] T. J. R Hughes, G. R. Feijoo, L. Mazzei, and J.-B. Quincy. The variational multiscale method – a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Eng.*, 166:3–24, 1998.

[82] T. J. R. Hughes, L. P. Franca, and M. Balestra. A new finite element formulation for computational fluid dynamics, V: Circumvent the Babuška–Brezzi condition: A stable Petrov–Galerkin formulation for the Stokes problem accommodating equal order interpolation. *Comput. Methods Appl. Mech. Eng.*, 59:89–99, 1986.

[83] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method.* Cambridge University Press, Cambridge, 1987.

[84] C. Johnson. Adaptive finite element methods for diffusion and convection problems. *Comput. Methods Appl. Mech. Eng.*, 82:301–322, 1990.

[85] C. Johnson. Adaptive finite element methods for the obstacle problem. *Math. Models Meth. Appl. Sci.*, 2:483–487, 1992.

[86] C. Johnson. Discontinuous Galerkin finite element methods for second order hyperbolic problems. *Comput. Methods Appl. Mech. Eng.*, 107:117–129, 1993.

[87] C. Johnson. A new paradigm for adaptive finite element methods. In J. Whiteman, ed., *Proc. MAFELAP 93*. John Wiley, 1993.

[88] C. Johnson and P. Hansbo. Adaptive finite element methods in computational mechanics. *Comput. Methods Appl. Mech. Eng.*, 101:143–181, 1992.

[89] C. Johnson and P. Hansbo. Adaptive finite element methods for small strain elasto-plasticity. *Finite Inelastic Deformations - Theory and Applications* (D. Besdo and E. Stein, eds), pp. 273–288, Springer, Berlin, 1992.

[90] C. Johnson and R. Rannacher. On error control in CFD. Proc. Int. Workshop *Numerical Methods for the Navier-Stokes Equations* (F.-K. Hebeker *et al.*, eds), pp. 133–144, *vol. 47 of Notes Num. Fluid Mech*, Vieweg, Braunschweig, 1994.

[91] C. Johnson, R. Rannacher, and M. Boman. Numerics and hydrodynamic stability: Towards error control in CFD. *SIAM J. Numer. Anal.*, 32:1058–1079, 1995.

[92] C. Johnson and A. Szepessy. Adaptive finite element methods for conservation laws based on a posteriori error estimates. *Comm. Pure Appl. Math.*, 48:199–234, 1995.

[93] G. Kanschat. Parallel and Adaptive Galerkin Methods for Radiative Transfer Problems. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 1996.

[94] G. Kanschat. Solution of multi-dimensional radiative transfer problems on parallel computers. *Parallel Solution of Partial Differential Equations* (P. Bjørstad and M. Luskin, eds), pp. 85–96, vol. 120 of *IMA Volumes in Mathematics and its Applications*, New York, 2000. Springer.

[95] G. Kanschat and R. Rannacher. Local error analysis of the interior penalty discontinuous Galerkin method. Preprint, Universität Heidelberg, 2002.

[96] H. Kapp. Adaptive Finite Element Methods for Optimization in Partial Differential Equations. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 2000.

[97] G. Kunert. An a posteriori residual error estimator for the finite element method on anisotropic tetrahedral meshes. *Numer. Math.*, 86:471–490, 2000.

[98] G. Kunert. A posteriori L_2 error estimation on anisotropic tetrahedral finite element meshes. *IMA J. Numer. Anal.*, 21:503–523, 2001.

[99] G. Kunert. Edge residuals dominate a posteriori error estimates for linear finite element methods on anisotropic triangular and tetrahedral meshes. *Numer. Math.*, 86:283–303, 2000.

[100] P. Ladeveze and D. Leguillon. Error estimate procedure in the finite element method and applications. *SIAM J. Numer. Anal.*, 20:485–509, 1983.

[101] M. G. Larson. A posteriori and a priori error estimates for finite element approximations of selfadjoint eigenvalue problems. *SIAM J. Numer. Anal.*, 38:608–625, 2000.

[102] M. G. Larson and A. J. Niklasson. Adaptive multilevel finite element approximations of semilinear elliptic boundary value problems. *Numer. Math.*, 84:249–274, 1999.

[103] W. Liu and N. Yan. A posteriori error estimates for some model boundary control problems. *J. Comput. Appl- Math.*, 120:159–173, 2000.

[104] W. Liu and N. Yan. Local a posteriori error estimates for convex boundary control problems. Preprint, University of Kent, 2002.

[105] L. Machiels, A. T. Patera, and J. Peraire. Output bound approximation for partial differential equations; application to the incompressible Navier-Stokes equations. In S. Biringen, editor, *Industrial and Environmental Applications of Direct and Large Eddy Numerical Simulation.* Springer, Berlin Heidelberg New York, 1998.

[106] L. Machiels, J. Peraire, and A. T. Patera. A posteriori finite element output bounds for the incompressible Navier-Stokes equations: application to a natural convection problem. Technical Report 99-4, MIT FML, 1999.

[107] J. Nitsche. Ein Kriterium für die Quasi-Optimalität des Ritzschen Verfahrens. *Numer. Math.*, 11:346–348, 1968.

[108] C. Nystedt. A priori and a posteriori error estimates and adaptive finite element methods for a model eigenvalue problem. Technical Report Preprint NO 1995-05, Department of Mathematics, Chalmers University of Technology, 1995.

[109] J. T. Oden. Finite elements: an introduction. In *Handbook of Numerical Mathematics, Vol. II, Finite Element Methods (Part 1)* (P.G. Ciarlet and J.L. Lions, eds), pp. 3–15, North-Holland, Amsterdam. 1991.

[110] J. T. Oden and S. Prudhomme. On goal-oriented error estimation for elliptic problems: Application to the control of pointwise errors. *Comput. Methods Appl. Mech. Eng.*, 176:313–331, 1999.

[111] J. T. Oden and S. Prudhomme. Estimation of modeling error in computational mechanics. Preprint, TICAM, The University of Texas at Austin, 2002.

[112] J. T. Oden, W. Wu, and M. Ainsworth. An a posteriori error estimate for finite element approximations of the Navier–Stokes equations. *Comput. Methods Appl. Mech. Eng.*, 111:185–202, 1993.

[113] M. Paraschivoiu and A. T. Patera. Hierarchical duality approach to bounds for the outputs of partial differential equations. *Comput. Methods Appl. Mech. Eng.*, 158:389–407, 1998.

[114] R. Rannacher. Error control in finite element computations. *Proc. Summer School Error Control and Adaptivity in Scientific Computing* (H. Bulgak and C. Zenger, eds), pp. 247–278. Kluwer Academic Publishers, 1998.

[115] R. Rannacher. A posteriori error estimation in least-squares stabilized finite element schemes. *Comput. Methods Appl. Mech. Eng.*, 166:99–114, 1998.

[116] R. Rannacher. *Finite element methods for the incompressible Navier-Stokes equations.* Fundamental Directions in Mathematical Fluid Mechanics (G. P. Galdi, J. Heywood, R. Rannacher, eds), pp. 191–293, Birkhäuser, Basel-Boston-Berlin, 2000.

[117] R. Rannacher. Duality techniques for error estimation and mesh adaptation in finite element methods. *Adaptive Finite Elements in Linear and Nonlinear Solid and Structural Mechanics* (E. Stein, ed.), vol. 416 of CISM Courses and Lectures, Springer, 2002.

[118] R. Rannacher and F.-T. Suttmeier. A feed-back approach to error control in finite element methods: Application to linear elasticity. *Computational Mechanics*, 19:434–446, 1997.

[119] R. Rannacher and F.-T. Suttmeier. A posteriori error control in finite element methods via duality techniques: Application to perfect plasticity. *Computational Mechanics*, 21:123–133, 1998.

[120] R. Rannacher and F.-T. Suttmeier. A posteriori error estimation and mesh adaptation for finite element models in elasto-plasticity. *Comput. Methods Appl. Mech. Eng.*, 176:333–361, 1999.

[121] R. Rannacher and F.-T. Suttmeier. Error estimation and adaptive mesh design for FE models in elasto-plasticity. *Error-Controlled Adaptive FEMs in Solid Mechanics* (E. Stein, ed.), John Wiley, to appear.

[122] S. Richling, E. Meinköhn, N. Kryzhevoi, and G. Kanschat. Radiative transfer with finite elements I. Basic method and tests. *A&A*, 380:776–788, 2001.

[123] T. Richter. Funktionalorientierte Gitteroptimierung bei der Finite-Elemente-Approximation elliptischer Differentialgleichungen. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 2001.

[124] M. Schäfer and S. Turek. Benchmark computations of laminar flow around a cylinder. (With support by F. Durst, E. Krause and R. Rannacher). *Flow Simulation with High-Performance Computers II (E. H. Hirschel, ed.), pp. 547–566, DFG priority research program results 1993-1995*, vol. 52 of Notes Numer. Fluid Mech., Vieweg, Wiesbaden, 1996.

[125] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54:483–493, 1990.

[126] K. Siebert. An a posteriori error estimator for anisotropic refinement. *Numer. Math.*, 73:373–398, 1996.

[127] E. Stein and S. Ohnimus. Coupled model- and solution-adaptivity in the finite-element method. *Comput. Methods Appl. Mech. Eng.*, 150:327–350, 1997.

[128] F.-T. Suttmeier. Adaptive Finite Element Approximation of Problems in Elasto-Plasticity Theory. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 1996.

[129] F.-T. Suttmeier. An adaptive displacement/pressure finite element scheme for treating incompressibility effects in elasto-plastic materials. *Numer. Meth. Part. Diff. Equ.*, 17:369–382, 2001.

[130] R. Verfürth. A posteriori error estimates for nonlinear problems. *Numerical Methods for the Navier–Stokes Equations* (F.-K. Hebeker et al., eds), pp. 288–297, vol. 47 of *Notes Numer. Fluid Mech.*, Vieweg, Braunschweig, 1993.

[131] R. Verfürth. A posteriori error estimates for nonlinear problems. Finite element discretization of elliptic equations. *Math. Comp.*, 62:445–475, 1994.

[132] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques.* Wiley/Teubner, New York Stuttgart, 1996.

[133] R. Verfürth. A posteriori error estimation techniques for nonlinear elliptic and parabolic pde's. *Rev. Eur. Élém. Finis*, 9:377–402, 2000.

[134] B. Vexler. A posteriori Fehlerschätzung und Gitteradaption bei Finite-Elemente-Approximationen nichtlinearer elliptischer Differentialgleichungen. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 2000.

[135] C. Waguet. Adaptive Finite Element Computation of Chemical Flow Reactors. Dissertation, Institute of Applied Mathematics, University of Heidelberg, 2000.

[136] R. Zamni. Integrationsfehleranalyse bei der adaptiven Finite-Elemente-Methode. Diploma thesis, Institute of Applied Mathematics, University of Heidelberg, 2001.

[137] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Numer. Meth. Eng.*, 24, 1987.

[138] M. Zlámal. On the finite element method. *Numer. Math.*, 12:394–409, 1968.

Index